

STIC-Biotech/ChemLib

80322

From: Portner, Ginny
Sent: Friday, November 15, 2002 3:54 PM
To: STIC-Biotech/ChemLib
Subject: 09/881,752

Please search and oligomer search SEQ ID No 212.

Ginny Portner
CM1, Art Unit 1645
Room 7e13
Mail box 7e12
(703) 308-7543

Edward Hart
Technical Info. Specialist
STIC/Biotech
CMI 6B02 Tel: 305-9203

RECEIVED
NOV 15 2002
STIC/Biotech Division
(STIC)

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 11/18/02
Date Completed: 11/18/02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:

NA Sequences: _____
AA Sequences: ☒ _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: ☒ _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into to two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, **.rnpn** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, **.rapn** and **.rapn**

The Pending database search results should not be left in the case because they contain data that is confidential.

THIS PAGE BLANK (USPTO)

According to the Pre Publication Rules, every patent application received by the United States Patent and Trademark Office after November 29, 2000 will be pre-published at eighteen months from the effective filing date. When the application is published the contents, including the sequences, will become prior art.

Two new databases have been created to hold the pre-published sequences:

Published_Applications_NA contains nucleic acid sequences; the search results will have the extension .rnpb.

Published_Applications_AA contains amino acid sequences; the search results will have the extension .rapb.

Each pre-published application is given a unique Publication Number. An example of a Publication Number is US20021234567A1. The "US" indicates the application was a U.S. application. The first 4 digits show the calendar year the application was published. The next 7 digits represent when the application was published. This 7-digit number starts at zero at the beginning of each calendar year. Each application published is given the next number in order. The "A" indicates a utility patent application and the "1" shows that this was the first time the application had been published. If the applicants submit changes to the application, they may request that the changed application be published again. In such instances, the "1" at the end of the number would be replaced by a "2".

Sequences in the PGPub database are public information; it is permissible to leave these results in the case.

THIS PAGE BLANK (USPTO)

Db 416 AKARLPDNRVLYEP 432

RESULT 14

ID Q8YM16 PRELIMINARY; PRT; 426 AA.

AC Q8YM16; 01-MAR-2002 (T-EMBLrel. 20, Created)
 DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
 DE Processing protease.
 GN ALR5125.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 NCBI_TaxId=103690;

RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840; Kuritz T., Sasamoto S.,
 RA Kaneo T., Nakamura Y., Wolk C.P.,
 RA Matanabe A., Itiguchi M., Ishikawa A., Kawashima K., Klmura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shlimpo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 "Complete genome sequence of the filamentous nitrogen-fixing
 Cyanobacterium Anabaena sp. strain PCC 7120.";
 RL DNA Res. 8:205-213(2001).
 DR EMBL; AF003598; BAB76824.1; -
 DR InterPro: IPR001431; Peptidase_M16.
 DR Pfam: PF00675; Peptidase_M16; 1.
 DR PROSITE: PS00143; INSULINASE; 1.
 KW Protease; Complete proteome.
 SQ SQUIDNCE 426 AA; 48051 MW; 7449B295CCBF5D0F CRC64;

Query Match 17.8%; Score 403; DB 16; Length 426;
 Best Local Similarity 27.0%; Pred. No. 2e-18;
 Matches 113; Conservative 96; Mismatches 189; Indels 20; Gaps 9;

QY 37 TLKNGLOV--SVPLENKTGVIEVDLYKYSGSRNETMGSGIAHMLHNFKSTKNKAG 94
 Db 18 TLBNGLITVEQMPVE---AVNLSTLMDYSSVESDAINGMAHFLHMLFKGTERRLASG 73
 QY 95 EPRKIYKREGVSNASTSPDITRFYFKTSQANDKSLSELFETMGSLNKEDEFLEPEROV 154
 Db 74 EPRHIEREAVVNAATSOQYTHYINTAFQDAKLAFLQIDVNLNASTIDEAFERERFV 133
 QY 155 VAEERRRTNSPIGMILYFEFNTAYVYHPTPIGFMDIONMTLKDKKHSLSLYOP 214
 Db 134 VLEIRK-RSEDNPRRTFRAMETAFAQLPYPYRRPVLGPESITISOLTPQMRDPLASWYOP 192
 QY 215 KNAIVLVGDVNSQKVELSKKHESLKNLDEKAIPTPYM--KEPKODGARTAVVHK--- 269
 Db 193 QSTAVAVVGLPEBOLLETFTEGFGNOKTTPHSPILPLHLEP---AFTELVREFFV 248
 QY 270 -DGVHLEWALGKVPFAFKKDOV--ALDALSRLLGSGKSSWLSQSELVDKRLASQAFSHN 327
 Db 249 DESIQOQARLLIMVVRVPGDLNLEQTYGLDVLGILAHGRTSRLVODLKEERGLVTSISVSN 308
 QY 328 MOLDSEVFLEFIAGNNVNAKALOKETIVALLLEKKKGELTQADLTKLNQKADFLSNL 387
 Db 309 MSNRLOQTF-YISAKCAVEDLAIVEEALIAQHIRKLOTELTETKEIAVVRKRVANRFLFGN 367
 QY 388 ESSSDVAGLEFADY-LVYONDIOGLTDYOROPLDKVSPLVAVANEYFKDQSTVYFLKP 444
 Db 368 ETPSDRAGLGYFGSLVGLDLEPAFNPYPAHTQTOGEAPDLLLANQYLCPEAYGVVYMKP 425

RESULT 15
 ID 067308 PRELIMINARY; PRT; 433 AA.
 AC 067308;
 DT 01-AUG-1998 (T-EMBLrel. 07, Created)
 DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)

DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
 DE Processing protease.
 GN MPP OR AQ.1271.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
 NCBI_TaxId=63363;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-VF5;
 RX MEDLINE=98196666; PubMed=9537320;

RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Anlay M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 "The complete genome of the hyperthermophilic bacterium Aquifex
 RT aeolicus.";
 RL Nature 392:353-358(1998).
 DR EMBL; AB000732; AAC07272.1; -
 DR InterPro: IPR001431; Peptidase_M16.
 DR Pfam: PF00675; Peptidase_M16; 1.
 DR PROSITE: PS00143; INSULINASE; 1.
 KW Protease; Complete proteome.
 SQ SQUIDNCE 433 AA; 51161 MW; 6C9E7AA227910B28 CRC64;

Query Match 17.8%; Score 403; DB 16; Length 433;
 Best Local Similarity 27.9%; Pred. No. 2.1e-18;
 Matches 119; Conservative 87; Mismatches 175; Indels 46; Gaps 12;

QY 38 LKNGLOVSVPLENKTGVIEVDLYKYSGSRNETMGSGIAHMLHNFKSTKNKAGFPD 97
 Db 26 LPMGAKLIVPRDD-TEAVLHWFVRKYSYEKDEKGMHFLHMLFNGTETKKGTEID 84
 QY 98 KIYKREGVSNASTSPDITRFYFKTSQANDKSLSELFETMGSLNKEDEFLEPEROVAE 157
 Db 85 RIIESLGINAGTSKDYTYVHVEIAHPYKQALEVLYLTMRKATLDEEMIEKEKPIVIE 144
 QY 158 EPRKRTNSPIGMILYFEFNTAYVYHPTPIGFMDIONMTLKDKKHSLSLYOPKNA 217
 Db 145 ELRGNKDN-PTVLMEEFEKLVYKVSYPYRPIIGFETIRKFTREKLKFKSYOPRNM 203
 QY 218 IYLVVGDVNSQKVELSKKH--ESLKNLDEKAIPTPYMKEPKODGARTAVVHKDGVHLE 275
 Db 204 AVVIVGVNPRKEVEEYMKTFEGEGRVPRKVOIPT---EPDQIGIRFKKLDPRIEKA 259
 QY 276 WVALGYVPAPFKHKDQYALDALSRLLGEGKSSWLSQSELVDKRLASQAFSHNMOLDSEV 335
 Db 260 YWIIIGWRVPAIGTDYGLVSEILLOGGRISVYFRLRKGLVYSYSGC-DNGRPDNI 318
 QY 336 FLPIAGNPVNAKALOKETIVALLLEKKKGELTQADLTKL---INOKADFLSNLESSSD 392
 Db 319 FIITATPEPPE-NEYKVKRVEFLIKETYE-NLDEQVEEAKSRILNSR----- 364
 QY 393 VAGFADLYOND-----IOGLTDYOROPLDKVS-----DLVRYANEYFKDQTS 437
 Db 365 ---LFEERVENDAFDIGITYVVRDLYF--RFDKNLSRVRRVARDWILFERTIKEDKY 419
 QY 438 TTVFLKP 444
 Db 420 SETLMWP 426

Search completed: November 18, 2002, 11:58:38
 Job time : 39 secs

OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 OC Ralstonia.
 ON NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salenouat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Arlet M., Billault A., Brottier P., Camus J.C., Cattolico L.,
 RA Chandler M., Choise N., Claudel-Bernard C., Cunac S., Demange N.,
 RA Gaspin C., Layle M., Moisan A., Robert C., Saurin W., Schlex T.,
 RA Siglier P., Thebaud P., Whalen M., Wincker P., Levy M.,
 RA Weissenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
 RL Nature 415:497-502(2002).
 DR EMBL: AL646059; CAD13915.1;
 DR InterPro: IPR001431; Peptidase_M16.
 DR Pfam: PF00675; Peptidase_M16; 1.
 DR PROSITE: PS00143; INSULINASE; 1.
 KW Hydrolase; Complete proteome.
 SO SEQUENCE 501 AA; 55423 MW; F04C814C6090D766 CRC64;

Query Match 18.7%; Score 425; DB 16; Length 501;
 Best Local Similarity 30.4%; Pred. No. 9.4e-20;
 Matches 133; Conservative 81; Mismatches 179; Indels 44; Gaps 15;

QY 33 HESVTLKNOLOYVSPLEKKTGVEVDLYLVKVSRRNETMGSKSIAMLEHNLKSKTKNKK 92
 DB 61 HE-YRLANLRLI-VAEDHRAFLPAHVMVYHAGSIDENHGTGVALHEHMKKTRAVG 118
 QY 93 AGEFDKIVRFSGVSNASTSPDITRYFIKTSQANLDKSLFAETMGSLNKEDEFLPER 152
 DB 119 PEFSRVAALGGRKAMTTRDTFTWFOGIEKSLADVALEDRANQLDKEKPEK 178
 QY 153 QVVAEERRRRRTNSPIGLYFRFENTAYYHRYHMTPIGFMDIQWTLKIKKHSLLY 212
 DB 179 NVYKERRRRRIDDSARATYEEOMLAVLFNAAPRYRPTIGPSSLDIMTQVODADQWYHKY 238
 QY 213 QPKNATLVVGVNSKVELESKKHESLKNDEKAIIPRYMK-EPKOGARAVYVHKOG 271
 DB 239 APNNATVVTGVDNDEVRROKORTYGLQ---PHALPRRYADDEKQVGVKTIW-KAP 294
 QY 272 VHELEWVALGYKVPAA---KHKDOVALDALSRLLGEGKSSWLOSELY---DKK-RLASQ 322
 DB 295 AENPVYVALYKAPRLDVEKDYDPRALYEVLSAVLDGYDANRLPNLYKRGDEKGLAAD 354
 QY 323 AFSHNQLODESVLEFIAGGNP---NVKALQKETVALLEKIKKEITQAEIDKLKINO 379
 DB 355 V-NAGYDGNRRGPSIFLLDGVADGHTTAE-IEQALRAQIDRIARDGVTEAEELKRVKQV 412
 QY 380 KADEISNLSES-----SSDYAGLF---ADLYVONDIOGLTDYQROQLDKVSLVAV 427
 DB 413 VAAQYIKRQSVFGQGEIGMAEMETGLSWRDLRIIEK-IKSVTPAVQO-----QV 461
 QY 428 ANEYFKDTQSTTVFLKP 444
 DB 462 AKTYFTEDNLVYATILP 478

RESULT 13
 OS0511 PRELIMINARY; PRT; 450 AA.
 AC OS0511;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Zinc protease.
 GN SC05837 OR SC9B10.04.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
 ON NCBI_TaxID=1902;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Oliver K., Harris D.;
 RL Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Parhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (Nov-1997) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kleser H.M., Denapalze D., Eichner A., Cullum J.,
 RA Kinsahl H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2)";
 RL Nature 417:141-147(2002).
 DR EMBL: AL009204; CA15794.1;
 DR InterPro: IPR001431; Peptidase_M16.
 DR Pfam: PF00675; Peptidase_M16; 1.
 KW Protease.
 SO SEQUENCE 450 AA; 49045 MW; 976D90CBB7F90F5 CRC64;

Query Match 18.2%; Score 412.5; DB 16; Length 450;
 Best Local Similarity 29.5%; Pred. No. 5.3e-19;
 Matches 129; Conservative 76; Mismatches 175; Indels 57; Gaps 11;

QY 38 LKNGLOVSPLEKKTGVEVDLYLVKVSRRNETMGSGIAHLEHNEKSTNKAERGE 97
 DB 23 LANKLRVY-LSBDHLPVAAVOCIMVDVSGRHEKGTGLAHLFELHMFQSGAQVGNHGF 81
 QY 98 KIVKRFQVSNASTSPDITRYFIKTSQANLDKSLFAETMGSL-NLKEDEFLPERQV 155
 DB 82 ELVQAGGSLNGTTSFERINRYFETMPAHQDELALMEADRMGSLAALADESEMENRDVY 141
 QY 156 AEBRWRTDNPISGLYFRFENTAYY---HRYHMTPIGFMDIQWTLKIKKHSLLY 212
 DB 142 KNERQRQYDNPVYGTAFELK---TALAYPRGHYHNPTRISMADELDAATLEDAARAFRTY 199
 QY 213 QPKNATLVVGVNSKVELESKKHESLKNDEKAIIPRYMKERPKOGARAVY----- 266
 DB 200 APNNATVVTGVDNDEVRROKORTYGLQ-----KOPRDGALDVGGEELR 251
 QY 267 -VHKDGVHLEWVALGYKVPAAFKHKDOVALDALSRLLGEGKSSWLOSELYDKKRLASQAFS 325
 DB 252 EYVEEVPARALMAATRYLPEDGTACDADALALTYGGESSRLNRYLRRTAVAAAGF 311
 DB 312 GLRL-----AGASLGLMDVYKTSQVDEVPYETAIIDELARFADECPTEEMER 361
 QY 368 TQAEIDKLKINQKADPISNLSESSDYAGLFADLYVONDIOGLTDYQROQLDKVSLVAV 427
 DB 362 AQAOLEREWLDRIQVYAGRADELCRYAVLFGDPQ-----ALTAVQR-VLEVTAEVDEY 415
 QY 428 ANEYFKDTQSTTVFLKP 444

QY 402 VQNDIGLTDYOROFDLKVSILVRANEXFKDTOSTTVELK 444
Db 419 IDDDLEALK-----AVTPDDIOKARFYFTPSRLTLAQLP 454

RESULT 10

ID 09KUG7 PRELIMINARY; PRT; 952 AA.
AC 09KUG7;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Protease, insulinase family/Protease, insulinase family.
GN VC0554.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L., Gill S.R., Nelson K.E., Read T.D., Tectelin H., Richardson D., Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P., McDonald L., Utterback T., Fleischmann R.D., Nieman W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C., Fraser C.M.;
RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae";
RT choleracae;
RL Nature 406:477-483(2000).
DR EMBL: AE004141; AAF93722.1; -;
DR TIGR: VC0554; -;
DR InterPro: IPR000794; Ketoacyl-synt.
DR InterPro: IPR001431; Peptidase_M16.
DR Pfam: PF00675; Peptidase_M16; 1.
DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_1.
DR Protease; Complete proteome.
SQ SEQUENCE 952 AA; 106354 MW; 47572DF6F5943137 CRC64;

Query Match

Best Local Similarity 19.0%; Score 432; DB 16; Length 952;
Matches 129; Conservative 81; Mismatches 182; Indels 60; Gaps 11;

QY 9 LLAGSSVLY---TLGASMAHASTLPK-----HESVTLKNGLOVVS 47
Db 7 LRFESLLLAGCCSSDSSLPFSSLPKGVTLVEYKAEKGMIPYSKYRLDNGSLVILS 66
QY 48 PLENKGTGIVEDVLYKVGSRNETMKGSGIAHMLEHNFKSTKNLKAQEPDKYKRGVS 107
Db 67 P-DDSDPLVHDVYHNGSAREEIGKGFHFEHMFQSGKHVGQDQHRLITEAGSL 125
QY 108 NASTSPITTYFKITSQANDKSLFPAETMGS--NLKDEFLPEQVVAEERRMTDN 165
Db 126 NGTINRRTYFETVPANOLEKIMLEADRMGFLDAVDSQKFEIORDTYKNEAQNNDN 185
QY 166 SPITGLTFRFENTAYYV--HYHMTPIGFMDIOWTLKDKKFKHSYLQYKNAIVLVGD 224
Db 186 RPIGIMMERKGEALYRPGHYSWOTIGYSDLDKRVYNDLKAFELRYTGNNAVLITGGD 245
QY 225 VNSQVFELESKKHESL--KNLDEKALPTPYMKERPKDGARTAVVHKDVLHEWVALGKY 283
Db 246 LDVKQTLAWQYKFGSIPKRPDVADP---KQPARLSEDRFTLDEDRVQGPMLLIWPT 301
QY 284 PAKHKQVALDALSRLLGKSSWLOSELV--DKRLASQAFSHNMLODESFLTAGG 342
Db 302 QYWAEOVADALASALGSGNSNLLYQELVKTOKAVDAADFQDCAALACTFYVYVAPAS 361
QY 343 NPVKAALOKEIALKEKKEITQAEIDKLKINOKADFISLJESSVAGLFADYLV 402
Db 362 GAKGKLARLVQETLVQLEKKGQCVSARLEQIIGSEEAANVPALES-----408
QY 403 QNDIGLTDYOROFDLKVSILVRANEXFKD 434

Db 409 ---VKG-----KVSQI--AANOTFFD 424

RESULT 11

ID 049145 PRELIMINARY; PRT; 709 AA.
AC 049145;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Protease.
GN PQOE.
OS Methylobacterium extorquens.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Methylobacterium group; Methylobacterium.
OX NCBI_TaxID=408;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AM1;
RX MEDLINE=96178988; PubMed=8606199;
RA Springer A.L., Ramamoorthi R., Lidstrom M.E.;
RT "Characterization and nucleotide sequence of pqqE and pqqF in Methylobacterium extorquens AM1";
RL J. Bacteriol. 178:2154-2157(1996).
DR EMBL: LA3135; AAB00962.1; -;
DR InterPro: IPR001431; Peptidase_M16.
DR Pfam: PF00675; Peptidase_M16; 1.
DR PROSITE: PS00143; INSULINASE; 1.
SQ SEQUENCE 709 AA; 77127 MW; 41659464578F05B6 CRC64;

Query Match

Best Local Similarity 18.8%; Score 427; DB 2; Length 709;
Matches 123; Conservative 83; Mismatches 203; Indels 8; Gaps 6;

QY 31 PKHESVTLKNGLOVSVLENKGTGIVEDVLYKVGSRNETMKGSGIAHMLEHNFKSTKN 90
Db 69 PEVSAFVLQDNGLDVYVVP--DHRAPVATHWVYVNRGNSADPDIGOSGIAHFEHLMFKGTER 127
QY 91 LKAGEFDKIKYKRGVSNASTSFDIRFYFKITSQANDKSLFPAETMGSINKDEFLP 150
Db 128 HPGAFESKRAVSSIGGQENATSTYDYAVFQVARDHLSTMAEADRMGSLVDDAVVAP 187
QY 151 EROVVAERMRRTDINSPIGLYFRFENTAYYVHPYHMTPIGFMDIOWTLKDKKFKHS 210
Db 188 EROVVAERMRRTDINSPIGLYFRFENTAYYVHPYHMTPIGFMDIOWTLKDKKFKHS 247
QY 211 YQPKNAIVLVGDVNSQVFELESKKHESLKNLDEKALPTPYMKERPKDGARTAVVHKD 270
Db 248 FYTPENAILVAGDVPEDEVRRLAEDYGRVTPQGAAPRLT--RPREPEPRAMRIAVADP 306
QY 271 GVHLEWVALGYKVA---FHKQVVALDALSRLLGKSSWLOSELVDKRLASQAFSHN 327
Db 307 KVEQPTLRVLYTPSCFTADGEGYALAEVVGSGSTFLRKVLLEGVVAVNGAWY 366
QY 328 M-LODESVEFLTAGSNPNVKAALOKEIALKEKKEITQAEIDKLKINOKADFISN 386
Db 367 MGSAMDTRFAYVAVPESGTLLEAEHIDRVLRVPEALGARA--IERAKIRLMAETVYS 425
QY 387 LESSSDVAGLFADYL--VQNDIGLTDYOROFDLKVSILVRANEXFKDTOSTTVEL 442
Db 426 SDSQSLARIYGSALAIGETVEEVRPVEIEATHTRLVAVAARYLVPARSVTGYL 482
RESULT 12
ID 08Y2E9 PRELIMINARY; PRT; 501 AA.
AC 08Y2E9;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative zinc protease protein (EC 3.4.99.-).
RSC0387 OR RS03355.

[illegible]

RESULT	8
ID	Q983M9
NAME	Q983M9 PRELIMINARY; PRT; 462 AA.
AC	Q983M9
DT	01-OCT-2001 (TREMBlrel, 18, Created)
DR	01-OCT-2001 (TREMBlrel, 18, Last sequence update)
DE	01-Oct-2001 (TREMBlrel, 18, Last annotation update)
GN	Protease.
OS	MNR825.
OC	Rhizobium loti (Mesorhizobium loti).
CC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OX	Phyllobacteriaceae; Mesorhizobium.
NCBI_TaxID=381;	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=MAFF303099;
RX	MEDLINE=21082930; PubMed=11214968;
RA	Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA	Watanabe A., Itozawa K., Ishikawa A., Kawashima K., Kimura T.,
RA	Kishida Y., KiyoKawa C., Kohara M., Matsuno M., Matsuo A.,
RA	Mochituki Y., Nakayama S., Nakazaki N., Shlimpo S., Sugimoto M.,
RA	Tateuchi C., Yamada M., Tabata S.;
RT	"Complete genome structure of the nitrogen-fixing symbiotic bacterium
RL	Mesorhizobium loti.";
RNA Res.	7:331-338(2000).
EMBL:	AP003013; BAB5381.1; "
DR	InterPro: IPRO01431; Peptidase.M16.
Pfam:	PF00675; Peptidase_M16; 1.
KW	Complete proteome.
SO	SEQUENCE 462 AA; 50952 MW; 1BC7B3DDFDC28C4 CRC64;

	Query Match	20.9%	Score 475	DB 16	Length 462
	Best Local Similarity	29.1%	Pred. No. 4.6e-23		
	Matches 134	Conservative 88	Mismatches 212	Indels 26	Gaps 8
QY	6 VKRLIGLSSVLLVT-----LGASMAQSYLPKHESVT--LKNIQLQVSVYLENKK	52			
Db	2 ISRAEMLRTLLATLSLAFTETGPVLADSTALPDKPIPAEFKVIDELLDNOMEVYVIP-DHR	60			
QY	53 TGVLEVNDLVKVGSKNNEIMGSGIAHMLEHLNFKSKNKLKAGEFDKIVKREGVSNASTS	112			
Db	61 APIVTHMWYKIKGSADPEPPKSGIAHFEHLMFKATITNNAAGEFDPAVSDIGSSNNAFTS	120			
QY	113 FDIETVFKTSOANDKLSELFAETMGSINLKEDEFLPEROYAVERRRRTNSPIGMXY	172			
Db	121 YDTYAFHETVAPSALEQMGFEADRRKNILITDDYIKTERDYVILERRSRIDNNPQAVLD	180			
QY	173 FRFENTAYVHYHPWTPIGFMDDIONWTLKDIKKHSHSLXYOPKMAIVLVGVDVNSQKVE	232			
Db	181 EYVDATLWQNDYRIPVIGMMQEMQOLNRDTAVAVYDKYIRRRNNAVLIYAGVEPPTVAA	240			
QY	233 LSKKTFEESI---KNIDEKALPTPYKKEPKQDGAFLAVVHKDGVHLIEWVALGKVPAAE---	286			
Db	241 LAEKTYGVVARGPDLPRPIRPV---EFEQNKRTVTYTLDAFVSVPSEFTQWVVPSTYHTA	296			
QY	287 KHKDQVADLALSRIIGBEKSSWLSQSEIVLDDKRLKASQAFSH-NMQLODESVPFLFIAGGN	345			
Db	297 KPGEEALDILAEIITGGNSRSLYQALVVKQSIASNAGVYFOGTMIDNITFVYGAQPCD	356			

OY 346 VKAAALQEIYAALLEKTKKEGIIETAOEDKIKINKAPFINLESSSVAGLADLYVO -N 404
 :
 Db 357 AKLDVAEAYDAVARIAASGVTFDELEKKAKDRITRSMVFARQDQSMATFYSTLATGC 416
 OY 405 DIQGLTIDYQRFLDKVSLDVLRANEFKFOGSTTVFLKP 444
 :
 Db 417 NVQDDQQMPDRIRKRYTADVEAAYAVARTLVLTARSGTLLP 456

	RESULT	9	
Q916C2	ID	PRLIMINARY;	PRT; 465 AA.
AC Q916C2;			
DT 01-MAR-2001 (TREMBLrel_16,			Created)
DT 01-MAR-2001 (TREMBLrel_16,			Last sequence update)
DT 01-JUN-2002 (TREMBLrel_21,			Last annotation update)
DE Probable zinc protease.			
GN PAO372.			
OS Pseudomonas aeruginosa.			
OC Bacteria; Proteobacteria; gamma			subdivision; Pseudomonadaceae;
OX Pseudomonas.			
OX NCBI_TaxID=287;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 15692 / PAOI;			
KX MEDLINE=20437337; PubMed=10964043;			
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,			
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,			
RA Gabor R.L., Goltzy L., Tolentino E., Westbrock-Wagman S., Yuan Y.,			
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,			
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,			
RA Reizer J., Salier M.H., Hancock R.E.W., Lory S., Olson M.V.,			
RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an			
RT opportunistic pathogen.";			
RL Nature 406:959-964(2000).			
DR EMBL: AE004475; AAC03761.1; -;			
DR InterPro: IPR001431; Peptidase_M16.			
DR Pfam: PF00675; Peptidase_M16.1.			
DR Complete proteome.			
QC SEQUENCE 465 AA; 51975 MW; 8D20E91407605A91 CRC64;			

Query Match	19.7%	Score 447	DB 16	Length 465
Best Local Similarity	29.4%	Pred. No. 3	16-21	
Matches 136	Conservative	96	Mismatches 183	Indels 48
				Gaps 15

QY	7	KRLGLSLVLTVTGASMAOS--YLPKHSVYTLKNGLOVSVPLENKTGVTEVDLVKYV	64
		: : : : : : : : : : : : : : : : : : : :	
Db	15	RRRVLT--LTLASLCLPLFAOAEATQPIHE-FSLNDGLKVI-VREDHRRPVPVVSOLMYRI	69
QY	65	GSRNTEKSGIAHLEHLNFKSTKNLKAEGEDKIVKRGCVSNASTSDIRYFLIKTRQ	124
		: : : : : : : : : : : : : : : : : : : :	
Db	70	GSSTYTPELTGLSHLLEHMFEGSKLKGEGESRVLRDGLGAENKFTTDYATAYQVYLAK	129
QY	125	ANLDESLFAETMGSNLMLKEDEFYPEROVAAEEERWRTDNPISGLMYFPFNATAYVNH	184
		: : : : : : : : : : : : : : : : : : : :	
Db	130	DLRLVALEMEADRYMHNLSLPYDQFSEIVEIVIEEERRLRTDQNPMLAFERFKAAYPASG	189
QY	185	YHWPYIGMDLIQNTLTLDKIKFHSYLYQPKALYLVYGDVNSQVFEELSKNHESL--K	242
		: : : : : : : : : : : : : : : : : : : :	
Db	190	YHPIPIGMADLQRTITDILRHMYESWVAAPNNATLVVYGDVADAEVKTLLAKRYFGEIPIWR	249
QY	243	NLDEKAIPTPMKERTKOGARTAUVNKHQGVNLEWALGKYVAF-----KHKOVALDAL	298
		: : : : : : : : : : : : : : : : : : : :	
Db	250	QLPRAKLE-LAEGERRLKIY-----RTQLPNIIMGFNPFSLASSEPRPVNMLRLIG	304
QY	299	RLLEGKSSWLOSELVDKRRLLASQAFSHNMLODESVFLTAGGNPNV-KAALQKEIYA	357
		: : : : : : : : : : : : : : : : : : : :	
Db	305	ALIDOGYSARLASRLERGEELVAGASTY-YDAFNNGDSLEFLVATNPVQKGTLLPQVEAG	363
QY	358	L--LEKTKKEITQALDELDKL-----INOKADPISULESSDYAGLFADYLL	401
		: : : : : : : : : : : : : : : : : : : :	
Db	364	LWKQDLDDKQMPSSAAELERVAQMIAGMYEKKDIAQASSITGQLES---VGL-SWKL	418

QY	273	HEWALGKGVAFKHKDDVALDALSLRLLGEEKSSWLOSELDDKKRLA--SOAFSGNNQL	330
Db	351	SQPTWLEGGYHRRPAMHPNDNAAYDIITASLSGRTSRKYSVEKRLNALNAGGSGFPED	410
QY	331	ODESVLEFLTAGNPNPVKAE---ALQKEIVALLERKLKKGEOIQAEDLKLINQKADFLSN	386
Db	411	KYPNIMLEFYALPAHNHTVDEVALATSKEL---DLKTEPVSAGELEERKTKQARGLINS	466
QY	387	LESSSDVAGLEPDIYVO-----NDIOGLDLYQROFLDKVSLDVRANFEKDT	435
Db	467	IDSNNMGMAQOLLEVEKTKGSMWRNLFKQOLDIVAVRP-----ADIQVAKAKATFTPE	516
QY	436	QST 438	
Db	517	NRT 519	
RESULT 6			
ID	09A2H7	PRELIMINARY:	PRT: 948 AA.
AC	09A2H7:		
CDT	01-JUN-2001 (TREMBLrel. 17, Created)		
FT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)		
PT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)		
DE	Peptidase, M16 family.		
GN	CC3584.		
OS	Caulobacter crescentus.		
OC	Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;		
NCBI_TaxID=155892;			
RP	SEQUENCE FROM N.A.		
RC	STRAIN-ATCC 19089 / CB15;		
RX	MEDLINE-21173698; PubMed-11259647;		
RA	Nieman W.C., Feldblyum T.V., Laud M.T., Paulsen I.T., Nelson K.E.,		
RA	Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,		
RA	Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,		
RA	Deboy R.T., Dodson R.J., Durkin A.S., Gwyn M.L., Haft D.H.,		
RA	Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,		
RA	Utterback T., Tran K., Wolf A., Yamahayan J., Emolaeva M., White O.,		
RA	Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;		
RT	"Complete genome sequence of <i>Caulobacter crescentus</i> ."		
RL	Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).		
DR	EMBL; AE006017; AAK2546.1; -		
DR	TIGR; CC3584; -		
DR	InterPro: IPR001431; Peptidase_M16.		
DR	Pfam; PF00675; Peptidase_M16; 2.		
KW	Complete proteome.		
SO	SEQUENCE 948 AA; 99142 MW; 281D91AFB4D13358 CRC64;		
Query Match			
	21.1%; Score 480; DB 16; Length 948;		
	Best Local Similarity 31.5%; Pred. No. 6.le-23;		
	Matches 133; Conservative 73; Mismatches 196; Indels 20; Gaps		
QY	33	HESVTLKNGIQVSVPLENKGTCGIVEDVLYUKGSRNETGKSGIAHMLHNLNFKSTKNLK	92
Db	45	YQQRVLANGMKVET-SRDTSTPNVSQVWYGSKDDPGRGSGFHLPHMLFKATRNMP	103
QY	93	AGEFPRKYKRGCGVSNASTSPDITRYFKITSOANDKSLFLFAETMGSINLKEDEFLPER	152
Db	104	NETVRLTEDVGGFNNASTWDDFTYTYEVVPAHNLERLIMAWADRLKSLIDEAVFASER	163
QY	153	OVADEBRHRTDNPISGLMYRFF-----NTPVYVHPHYWPIGMDQNTQMLDIKKE	207
Db	164	DYKKEELKQRLVADRYG-----RFFALSTPOGSFAVHPYORPGISITIEDLAATVDVRAF	219
QY	208	HSLYYQPKNAIVLVGDVNSQKFEFLSKKHFESLKNLDEKATPTPYMKPEKDGARTAV	267
Db	220	HRTYVRPDAAALIIYGNFGQTKLDAMIDKYFSSL-TTPGALIPKTVAVPARPTGKTYWT	278
QY	268	HNDGVHLEVALVGLGYRPAFKHDDQVALDALSLRLLGEEKSSWLOSELDDKKRLASQAFS--	325

```

Db      279 YGPNVPLDALATITWLAAPAAADKADAPALAVLDAIILTAGKSSRLYDLSVYDQKIAQSVSSA 338
Oy      326 -HNMOLODESVFLTAGGNPNVAKAPALQKEIYALLLEKKGITTOAEIDKLINQKADFI 384
Db      339 PNNAPPGLEFYVGAIMAGGTVTAQGAAPALRAQVA---RVNDGLVTPEALAEAKAGLADAV 395
Oy      385 SLESSSPVAGLEFADLYLVQ--NDIGLNDYQKQFIDLKSDIVRYANEXEKRTOGSTVYL 442
Db      396 RRRE-ELDGRGPAICATGATQTEGDAANAANSILAKQAVTADIQARAQYLIADRRVTINW 454
Oy      443 KP 444
Db      455 LP 456

RESULT 7
P74305 PRELIMINARY; PRT: 524 AA.
AC P74305.
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE PROTEASE.
GN POE OR SLLO915.
OC Synecchocystis sp. (strain PCC 6803).
OX Bacteria; Cyanobacteria; Chroococcales; Synecchocystis.
RN NCBI_TaxID=1148;
RP [1]
RX SEQUENCE FROM N.A.
MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nairu K., Okumura S.,
RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecchocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90914; BAA18399.1;
DR InterPro: IPR001431; Peptidase.M16.
DR Pfam: PF00675; Peptidase.M16; 1.
DR PROSITE; PS00143; INSULINASE; 1.
KW Protease; Complete proteome.
SQ SEQUENCE 524 AA; 59588 MW; 24824D8F31EBAD0E CRC64;

Query Match: 21.1%; Score 479; DB 16; Length 524;
Best Local Similarity 26.9%; Pred. No. 3e-23;
Matches 141; Conservative 88; Mismatches 186; Indels 110; Gaps 13.

Oy      5 SVKRLGLSSVLTVLGASM-----HAQSYLP-----KHESYTLKNG 41
Db      8 SLSRCLIVTALLVGHGIGLSMHPAIALADLATNNRNSGTLLFPIYRATQRTTERQDLNG 67
Oy      42 LQVAVPLENKTGVLEVDLVLYVGSRNETMGKSGIAHMLEHLNFKSTKN----- 90
Db      68 LKFI-VMENNENPARYVSFTYFDVGVDPEYGTGVAHLEHMAFGTERIGTKPFTQEQ 126
Oy      91 -----LKAEPFKIVARPGVS-N 108
Db      127 LLDLDQVFAQITTPARAGDKTGEOKLOEFQKQIDQAOADILKQNEFGQIOMAGVGYN 186
Oy      109 ASTSFDIRFYFIKTSQANLDKSLLEFAETMGSYLNKEDFLEPQVVAEERWRDINSPI 168
Db      187 AATSDADAFITFYSLEPSNKLLEIMLSLESEFF--LEVFREFRYQEVILEERKMRKENNPV 244
Oy      169 GMLYFRFENFAVYVPHYMTPIGFMDIDIONMTLKDICKFHSLSLYOPKNAIVLVGDSNQ 228
Db      245 GOMVEEFDITFAFTKPRYRRPVIGYDEDIRNLSRODVTDFEKKYIIPGNMTIIVGDKVD 304
Oy      229 KYVELSKHNFESLKLMDKCAIPTPRMK--EPKQDARFIAVYHKGDSVL-----EWALGY 281
Db      305 QVKSIAQYVGFGR---DQRPETPDQVTVLEPQO-----TOQKRNKINTLPSQPYEFGY 353

```

```
Db 64 AGEFDEIVGFGVDNASTGFDYTHYIICAKKNLDEKALEFAELMANLMLKDEEFOPER 123
QY 153 OVAEERRRRTDNSPIGMLYREFRNTAYVHPHWTPIGMDIOQMTLKDIKKFSLTY 212
Db 124 AYLLERRRTDNNPGLYLFLEFNHAFMHPHWTPIGFKDIENNSIDIKFHSITY 183
QY 213 OPRNATLVVGVNDSQVLELSSKHFESLKNLDEKAIPTPYMKPEKODGARTVVRKDV 272
Db 184 OPRNATLVVSGDLESKEVEFELSKHFEKTKN--TKTTPKHTHEKEPDGDKRYILKHS- 240
QY 273 HLEWVALGVKVPFKHKDOVALDLSRLLEGKSSWMLQSELYDKRLASQAFSHNQLOD 332
Db 241 DEFLALAVKIPNFKHKDIPALNALSELGSKSSIMSEILIDKMLINDYAYVNDICD 300
QY 333 ESEVLEIAGNPNVKAELQKEIVALLEKIKGELTQAEIDKIKNADEKINLESSSD 392
Db 301 ENLFIFICNCPNVNAKEKELKLTIDKLMKGISOKDQKRVKNKNSDIFSLNANA 360
QY 393 VAGLFADYVQNDIOGLTDYQROFLDKVSDLVANEXEKPQSTTFELK 443
Db 361 VANIVSYLARGDINPLNLENKDIQNLKDLISCAKKRYIQNSTVILR 411
```

```
RESULT 4
ID 08YB63 PRELIMINARY: PRT; 464 AA.
AC 08YB63:
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
GN BMEI1037.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_Taxid=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RA MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapral V., Redkar R.J., Patra G., Mijer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Rezik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,
RA Selkov E., Elser P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyriades N., Overbeek R.,
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RT Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL; AE009736; AAL54279.1; -.
DR InterPro: IPR001431; Peptidase_M16.
DR Pfam: PF00675; Peptidase_M16; 1.
DR PROSITE: PS00143; INSULINASE; 1.
KW Hydrolyase; Complete proteome.
SQ SEQUENCE 464 AA; 51266 MW; 21999CDOC072A533D CRC64;
```

Query Match 22.4%; Score 507.5; DB 16; Length 464;
Best Local Similarity 30.8%; Pred. No. 3.5e-25;
Matches 135; Conservative 85; Mismatches 194; Indels 25; Gaps 7;

```
QY 26 AOSYLPR-----HESVTLKGLQVSVPLENTGVLEVDVLYVGSRNEMTSGIAHM 79
Db 2 AQAALPEIRLSDGVSNFTLPNGQVVVIR-DHRAPVYVQWVYHVGADAPVSGIAHF 60
QY 80 LEHLNKRSTKNLKAAGEFDKIVKRFQVSNASTSFDTITRYIKTSQANLKSLELPAETMG 139
Db 61 LEHLNKRSTKNLKAAGEFDKIVKRFQVSNASTSFDTITRYIKTSQANLKSLELPAETMG 120
QY 140 SLNLKDEFLPERQVVAEERRRRTDNSPIGMLYREFRNTAYVHPHWTPIGMDIOQMT 199
Db 121 NLVLEDAVAKTEREVILLEERRRRTDNSPIGMLYREFRNTAYVHPHWTPIGMDIOQMT 180
QY 200 TLKDIKFFHSLYQPRNATLVVGVNDSQVLELSSKHFESLKNLDEKAIPTPYMKPEKQ 259
```

```
Db 181 SLKNADIEFYQNYVTNNATLVIVAGDVTPEVRBLAKMTAWANVHKRAE-VLLRRRPEPAK 239
QY 260 DGAARTAVVHNDGVHLEWVALGVKVPF-----KHKQVALLDLSRLLEGKSSWMLQ 310
Db 240 HAARVVTLDHERVSTPSFRISWLVPSIANEKRRANVPGDAPALDLSELIGSSQSLRLY 299
QY 311 SELVDKRLASQ-AFSHNMQLODESVFLTAGNPNVKAELQKEIVALLEKIKGELTQ 369
Db 300 QQLIVAGIAELGAYDGDALDDGTFSYGVPRNGASLGDEYKAAVQYDRILIRSGVYQ 359
QY 370 AEIDKIKNADEKIFISNESSDVAELFADYL---VQNDIOGLTDYQROFLDKVSDLY 425
Db 360 AEIDQARNRFLKAVIARSDQTMARIGSALSGVQTDIOKMPDLIK---SVTVDOIK 416
QY 426 RVANEXEKPQSTTFELK 444
Db 417 DVARRVLVQDAVTSYLLP 435
```

```
RESULT 5
ID 08YV4 PRELIMINARY: PRT; 528 AA.
AC 08YV4:
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
GN ADL1940.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_Taxid=103690;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsuno M., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genome sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RT DNA Res. 8:205-213(2001).
DR EMBL; AP003587; BAB73639.1; -.
DR InterPro: IPR001431; Peptidase_M16.
DR Pfam: PF00675; Peptidase_M16; 1.
DR PROSITE: PS00143; INSULINASE; 1.
KW Complete proteome.
SQ SEQUENCE 528 AA; 59607 MW; 324F21BEADD84E12 CRC64;
```

Query Match 21.4%; Score 486; DB 16; Length 528;
Best Local Similarity 30.4%; Pred. No. 1.1e-23;
Matches 147; Conservative 67; Mismatches 155; Indels 114; Gaps 14;

```
QY 38 LKNGLVSVPLENTGVLEVDVLYVGSRNEMTSGIAHMLDHLNFKST----- 88
Db 69 LDGGMFV-FLERHQAPVVSFLFYADVGVDEDDGTGVANFLEHLAFKTTTIGTENYQ 127
QY 89 -----KNLKAAGEFDKIVKRFQV 105
Db 128 AEKPLLERLQDLOIRAAKANGKODVYVRLQATPFKEVESQACKLVKQMLGIVDSQG 187
QY 106 VS-NASTSPDITRYIKTSQANLKSLELPAETMGSINLKE-----EFLPERQVY 155
Db 188 VGLNATSTATRYFSF-----SNKLELM-----MSLESDFLDVPIRREYKRDVY 237
QY 156 AEERRRRTDNSPIGMLYREFRNTAYVHPHWTPIGMDIOQMTLKDIKKFSLTYQPR 215
Db 238 LEERRRRTDNSPIGMLYREFRNTAYVHPHWTPIGMDIOQMTLKDIKKFSLTYQPR 297
QY 216 NALIVVGVNDSQVLELSSKHFESLKNLDEKAIPTPYMK---EPKODGARTVVRKDV 272
Db 298 NLTIAVGVDEVAQVRLAQTGYGRY-----KAAPKQSKIAATEPRQ--TOTREVLLEIA 350
```


Query Match 100.0%; Score 2270; DB 16; Length 444;
Best Local Similarity 100.0%; Pred. No. 3e-140;
Matches 444; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKHFSVRRLGLSSVLLVLTGASMAHQAOSYLPKHESVTLKNGLOVYVPLENKTGVEVDV 60
DB 1 MKHFSVRRLGLSSVLLVLTGASMAHQAOSYLPKHESVTLKNGLOVYVPLENKTGVEVDV 60
QY 61 LYKVSRRNEMTGKSGIAHMLNFKSTKNLKAGEPDKIVKRGVSNASTSPDITRYFI 120
DB 61 LYKVSRRNEMTGKSGIAHMLNFKSTKNLKAGEPDKIVKRGVSNASTSPDITRYFI 120
QY 121 KTSQANLDSLELFAETMGSINLKEDEFLPERQVAAEERRRRTDNSPIGMLYFFFPNTAY 180
DB 121 KTSQANLDSLELFAETMGSINLKEDEFLPERQVAAEERRRRTDNSPIGMLYFFFPNTAY 180
QY 181 VYHRYHMTPIGFMDDIQNMTLKDKIKFHSILYYQPKNAIVLVVGDVNSQKVFELSKKHES 240
DB 181 VYHRYHMTPIGFMDDIQNMTLKDKIKFHSILYYQPKNAIVLVVGDVNSQKVFELSKKHES 240
QY 241 LKNDEKAIPTPYMKERKODGARTAVVHKDGVHLEWALGKYVAFKHKDQVALDALSLRL 300
DB 241 LKNDEKAIPTPYMKERKODGARTAVVHKDGVHLEWALGKYVAFKHKDQVALDALSLRL 300
QY 301 LGECKSSWLOSELVDKRLASQAFSHNQLODESVFLFIAGNPNVKAALQKEIVALLE 360
DB 301 LGECKSSWLOSELVDKRLASQAFSHNQLODESVFLFIAGNPNVKAALQKEIVALLE 360
QY 361 KLKKGETQAEIDKIKINQKADFTISNLESSSDVAGLFADYLVQNDIGLTDYQOQFDLKL 420
DB 361 KLKKGETQAEIDKIKINQKADFTISNLESSSDVAGLFADYLVQNDIGLTDYQOQFDLKL 420
QY 421 VSDLVRAVANEYFKDOSTTYFLKP 444
DB 421 VSDLVRAVANEYFKDOSTTYFLKP 444

RESULT 2
Q92M12 PRELIMINARY; PRT; 443 AA.
AC Q92M12;
ID Q92M12;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative zinc protease.
GN JHP0411.

OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Dolg P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tumbino P.J., Carnuso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Kerberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
DR EMBL: AE001475; AAD05993.1; -;
DR InterPro: IPR001431; Peptidase_M16.
DR Pfam: PF00675; Peptidase_M16; 1.
DR PROSITE: PS00143; INSULINASE; 1.
KW Complete proteome.
SQ SEQUENCE 443 AA; 50032 MW; 861D38FC7E53FEB2 CRC64;

Query Match 94.7%; Score 2149.5; DB 16; Length 443;
Best Local Similarity 95.7%; Pred. No. 2.2e-112;
Matches 425; Conservative 7; Mismatches 11; Indels 1; Gaps 1;

QY 1 MKHFSVRRLGLSSVLLVLTGASMAHQAOSYLPKHESVTLKNGLOVYVPLENKTGVEVDV 60

DB 1 MKHFSVRRLGLSSVLLVLTGASMAHQAOSYLPKHESVTLKNGLOVYVPLENKTGVEVDV 60
QY 61 LYKVSRRNEMTGKSGIAHMLNFKSTKNLKAGEPDKIVKRGVSNASTSPDITRYFI 120
DB 61 LYKVSRRNEMTGKSGIAHMLNFKSTKNLKAGEPDKIVKRGVSNASTSPDITRYFI 120
QY 121 KTSQANLDSLELFAETMGSINLKEDEFLPERQVAAEERRRRTDNSPIGMLYFFFPNTAY 180
DB 121 KTSQANLDSLELFAETMGSINLKEDEFLPERQVAAEERRRRTDNSPIGMLYFFFPNTAY 180
QY 181 VYHRYHMTPIGFMDDIQNMTLKDKIKFHSILYYQPKNAIVLVVGDVNSQKVFELSKKHES 240
DB 181 VYHRYHMTPIGFMDDIQNMTLKDKIKFHSILYYQPKNAIVLVVGDVNSQKVFELSKKHES 240
QY 241 LKNDEKAIPTPYMKERKODGARTAVVHKDGVHLEWALGKYVAFKHKDQVALDALSLRL 300
DB 241 LKNDEKAIPTPYMKERKODGARTAVVHKDGVHLEWALGKYVAFKHKDQVALDALSLRL 300
QY 301 LGECKSSWLOSELVDKRLASQAFSHNQLODESVFLFIAGNPNVKAALQKEIVALLE 360
DB 301 LGECKSSWLOSELVDKRLASQAFSHNQLODESVFLFIAGNPNVKAALQKEIVALLE 360
QY 361 KLKKGETQAEIDKIKINQKADFTISNLESSSDVAGLFADYLVQNDIGLTDYQOQFDLKL 420
DB 361 KLKKGETQAEIDKIKINQKADFTISNLESSSDVAGLFADYLVQNDIGLTDYQOQFDLKL 420
QY 421 VSDLVRAVANEYFKDOSTTYFLKP 444
DB 421 VSDLVRAVANEYFKDOSTTYFLKP 444

RESULT 3
Q9PPB5 PRELIMINARY; PRT; 416 AA.
AC Q9PPB5;
ID Q9PPB5;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative zinc protease.
GN CJO805.

OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC SFRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Jorgels K., Karlyshev A.V., Moule S., Pallan M.J., Penn C.W.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
DR EMBL: AL139076; CAB73070.1; -;
DR InterPro: IPR001431; Peptidase_M16.
DR Pfam: PF00675; Peptidase_M16; 1.
DR PROSITE: PS00143; INSULINASE; 1.
KW Complete proteome.
SQ SEQUENCE 416 AA; 48043 MW; 9A40KC42050B0DCE CRC64;

Query Match 55.1%; Score 1251.5; DB 16; Length 416;
Best Local Similarity 57.2%; Pred. No. 8.2e-74;
Matches 235; Conservative 75; Mismatches 98; Indels 3; Gaps 2;

QY 33 HESVTLKNGLOVYVPLENKTGVEVDVLYKVSRRNEMTGKSGIAHMLNFKSTKNLK 92
DB 4 YEKQLNKLELVYALPYNKNSDVSVDFYKVSRRNEMTGKSGIAHMLNFKSTKNLK 63
QY 93 AGEFDKIVKRGVSNASTSPDITRYFIKTSQANLDSLELFAETMGSINLKEDEFLPER 152

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 11:56:19 ; Search time 35 Seconds

(without alignments)
2613.855 Million cell updates/sec

Title: US-09-881-752A-212

Perfect score: 2270
Sequence: 1 MKHFSVKRLGLSSVLAIVTL.....VRVANEYFKDQSTTVFLKRP 444

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2370	100.0	444	16	025656
2	2149.5	94.7	443	16	092M12
3	1251.5	55.1	416	16	09PBP5
4	507.5	22.4	464	16	08YB63
5	486	21.4	528	16	08YVNA
6	480	21.1	948	16	09A2H7
7	479	21.1	524	16	P74305
8	475	20.9	462	16	0983M9
9	447	19.7	465	16	0916C2
10	432	19.0	952	16	09XU67
11	427	18.8	709	2	049145
12	425	18.7	501	16	08Y2E9
13	412.5	18.2	450	16	050511
14	403	17.8	426	16	08YK16
15	403	17.8	433	16	067308
16	401.5	17.7	951	16	09RT29

17	372.5	16.4	414	16	08YTH4	08Yth4 anabaena sp
18	372.5	16.4	990	16	09P62	09P62 xylella fas
19	364.5	16.1	255	2	09ANJ1	09anj1 bradyrhizob
20	357.5	15.7	412	16	092IX7	092ix7 rickettsia
21	356.5	15.7	945	16	08YV31	08yv31 anabaena sp
22	354.5	15.6	927	16	09A579	09a579 caulobacter
23	343.5	15.1	430	16	P73670	P73670 synecocyst
24	341.5	15.0	428	16	05159	05159 synecocyst
25	335	14.8	408	16	08R653	08r653 fusobacteri
26	324.5	14.3	933	16	051486	051486 borrelia bu
27	323	14.2	534	8	09T259	09t259 solanum tub
28	319.5	14.1	300	2	09T1G1	09t1g1 rhodothermu
29	319	14.1	534	10	041445	041445 solanum tub
30	318.5	14.0	416	16	097GP6	097gp6 clostridium
31	310.5	13.7	414	16	092RP6	092rp6 rhizobium m
32	309.5	13.6	490	16	08YER9	08yer9 brucella me
33	307	13.5	434	16	08UH97	08uh97 agrobacteri
34	303	13.3	430	16	0985V3	0985v3 rhizobium l
35	295	13.0	528	10	09AXQ2	09axq2 cucumis mel
36	294.5	13.0	457	3	09P7X1	09p7x1 schizosacch
37	290	12.8	531	10	09SGA7	09sga7 arabidopsis
38	287	12.6	413	16	09KAB5	09kab5 bacillus ha
39	282.5	12.4	414	16	08XJ37	08xj37 clostridium
40	279	12.3	427	16	08YTH3	08yth3 anabaena sp
41	274.5	12.1	527	10	094KT0	094kt0 avicennia m
42	273	12.0	931	16	08XAV7	08xav7 escherichia
43	265	11.7	480	4	096DD2	096dd2 homo sapien
44	264.5	11.7	941	2	09XBW5	09xbw5 porphyromon
45	261	11.5	406	16	097H10	097h10 clostridium

ALIGNMENTS

RESULT 1

ID 025656 PRELIMINARY; PRT; 444 AA.

AC 025656;

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Protease (P00E).

GN HP1012.

OS Helicobacter pylori (Campylobacter pylori).

OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;

OC Helicobacter.

OX NCBI_TaxID=210;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=26695 / ATCC 700392;

RX MEDLINE=97394467; PubMed=9252185;

RA Tomb J.-F., White O., Kervavage A.R., Clayton R.A., Sutton G.G.,

RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,

RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,

RA Loftus B., Richardson D., Dodson R., Khaliq H.G., Glodek A.,

RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,

RA Berg D.E., Gocayne J.D., Uutterback T.R., Peterson J.D., Kelley J.M.,

RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,

RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,

RA Venter J.C.;

RA "The complete genome sequence of the gastric pathogen Helicobacter

RT pylori.";

RT Nature 388:539-547(1997).

RL EMBL: AE000609; AAD08056.1; -.

DR TIGR: HP1012; -.

DR InterPro: IPR001431; Peptidase_M16.

DR InterPro: IPR003880; Pantine_attach.

DR Pfam: PF00675; Peptidase_M16; 1.

DR PROSITE: PS00143; INSULINASE; 1.

DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_1.

KW Hypothetical protein; Protease; Complete proteome.

SQ SEQUENCE 444 AA; 50330 MW; 3A6092EF0792EAB2 CRC64;

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

Query Match          5.6%: Score 126; DB 3; Length 129;
Best Local Similarity 32.0%: Pred. No. 0.00019.
Matches 39; Conservative 25; Mismatches 50; Indels 8; Gaps 5.

OY 38 LKNGLOV--VSVPLEKTKGVLEVDLYKVGSHNETMGKSGIAHMLEHLNFKSTKNL-KAG 94
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 LANGIKVILMSPPPTDKSSA-ALDV--HIGSIDDPNINAGLSHFCEHMLFLCTKKKPKEN 57

OY 95 EFDKIVRRGGVSNASTSFIDITRYFIKTSQANLDKSLFELFAETMGSINLKEDEFLPEROV 154
   | : : : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 58 EYSQSLSEHAGSSNAFTGERTHNTYFDVSHHELEGALDRFAQFF--LCPLFDESCKDREV 115

OY 155 VA 156
   |
Db 116 NA 117

```

Db 354 GWNTLVGGKRGARGMFEIINVLTREGLHVEDIILHMFQY----OKLRAGPQEW 409
QY 353 -----KEIVALLKIKKGETTOAELDKIKINOKADFISSNESSDVG-----LF 397
Db 410 VFOECKDLNANAFRFKDKERPG-----YTSKIAKGKHYPLNGVLT 451
QY 398 ADYLVONDIQGLNDYORFQDLKVSDF---VRA 428
Db 452 AEVL-----LEEFPRDLIDWLDLRLPENRVA 479

RESULT 12

US-09-187-049-8
Sequence 8, Application us/09187049
Patent No. 6117666

GENERAL INFORMATION:

APPLICANT: Lampka, Gayle K.
TITLE OF INVENTION: PLASTID PROTEOLYTIC PROCESSING ENZYME
TITLE OF INVENTION: THAT CLEAVES PRECURSOR POLYPEPTIDES
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: BRINKS HOFER GILSON & LIONE
STREET: P.O. Box 10395
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60610
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/187,049
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/695,177
FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Martin, Alice O.
REGISTRATION NUMBER: 35,601
REFERENCE/DOCKET NUMBER: 7814/16
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312 321-4200
TELEFAX: 312 321-4299

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: MPP B Subunit from rat

US-09-187-049-8

Query Match 6.7%; Score 151; DB 3; Length 124;
Best Local Similarity 33.6%; Pred. No. 7,4e-07;

Matches 41; Conservative 20; Mismatches 57; Indels 4; Gaps 2;

QY 38 LKNGLOVSVPLENKGIVLVDLYVGRSNETMGRSGIAHMLEHNFSTKLNKAGEF 96
Db 1 LKNGLRVAVS---ENSGISTCTVGLIMIDAGSRVEKKNNGRAHFLFHNAAFQTKRSQDL 57
QY 97 DKIVKRGVSNASTSPDITRYFKTSQANLDSLELFAETMGSILNKEDEFLPEROVA 156
Db 58 ELEIENNGAHLNATYSNEQIVYAKAKSKDLPRAVELADIIONSTLGEAEIEREGRVIL 117
QY 157 EE 158
Db 118 RE 119

RESULT 13

US-09-187-049-6
Sequence 6, Application us/09187049
Patent No. 6117666

GENERAL INFORMATION:

APPLICANT: Lampka, Gayle K.
TITLE OF INVENTION: PLASTID PROTEOLYTIC PROCESSING ENZYME
TITLE OF INVENTION: THAT CLEAVES PRECURSOR POLYPEPTIDES
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: BRINKS HOFER GILSON & LIONE
STREET: P.O. Box 10395
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60610
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/187,049
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/695,177
FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Martin, Alice O.
REGISTRATION NUMBER: 35,601
REFERENCE/DOCKET NUMBER: 7814/16
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312 321-4200
TELEFAX: 312 321-4299

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: YDCC Swiss Protein P31828

US-09-187-049-6

Query Match 6.4%; Score 145.5; DB 3; Length 136;
Best Local Similarity 31.1%; Pred. No. 2,8e-06;

Matches 41; Conservative 23; Mismatches 59; Indels 9; Gaps 3;

QY 38 LKNGLOVSVPLENKGIVLVDLYVGRSNETMGRSGIAHMLEHNFSTK-----NLKA 93
Db 1 LDNGLRVAVS---ENSGISTCTVGLIMIDAGSRVEKKNNGRAHFLFHNAAFQTKRSQDL 60
QY 94 GEFDKIVKRGVSNASTSPDITRYFK---TSQANLDSLELFAETMGSILNKEDEFLP 150
Db 61 ETFESMGLRGRGVNATYSTDEVYQVSLPTQKONLDQVMAIFSEMSNATATEKLEVDA 120
QY 151 EROVVAEERRR 162
Db 121 ERGVITEE--WR 130
RESULT 14
US-09-187-049-7
Sequence 7, Application us/09187049
Patent No. 6117666
GENERAL INFORMATION:
APPLICANT: Lampka, Gayle K.
TITLE OF INVENTION: PLASTID PROTEOLYTIC PROCESSING ENZYME
TITLE OF INVENTION: THAT CLEAVES PRECURSOR POLYPEPTIDES
NUMBER OF SEQUENCES: 13

1

308 WLOSELVDKKRLASQAFSH---NMOLODES-----VFLFIAGNPVYKAERALO-- 352

Db 297 HLQL--YKIVPIKDIRNLVTPIPDLOQYIKSNPGHYLGLIGHEGPGSLSELSK 353

D5 297 HKQL--YKIVPKDIRNLVYTFPIPDLDQYYKSNPGHYLGHLIGHEGPGSLSELKSK 353

```

0Y      308 WLOS EVDK RKLASQAFSH---NMQLDDES-----VLF L IAGSNPNVKAELQ-- 352
          : || :: :: |           | : | : |           : | : |           : || |

```

STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: BRAINOT14
 CLONE: 1593490
 US-08-895-521-1

Query Match 10.3%; Score 233; DB 2; Length 489;
 Best Local Similarity 23.9%; Pred. No. 1e-13;
 Matches 99; Conservative 80; Mismatches 196; Indels 40; Gaps 14;

QY 7 KRLLGSSVLLV--TLGASMH-----AQSYL--PKHESVTLKNGLOVSVPLE 50
 DB 16 RRLMGFSESLIRGAAGRSILYFGENRLRSTQATQVYLVNPEIRYTCLESGLRVASE--D 73
 Y 51 NKTGVLEVVDLYKVGSRNETMGKSGIAHMLEHLENFKSTNKLKAGEPDKYKRGVSNAS 110
 DB 74 SGISTCTVGLMIDAGSRYEENKNGTAHLEHNAFGTKKRSOLDLELEIENNGAHLNAY 133
 QY 111 TSPDITRYFKITSQANLMDKSLFPAETMGSNLKEDFLPERQVAAEERWRKTDNSPIGM 170
 DB 134 TSEQTVYVYAKAFSKDLPRAVELADIIONSTLGEAEIEREKVYLIREMOEVEFN--LQE 191
 QY 171 LYFRFEN-TAYVYHPYHWPPIGFMDIIONMTLDIRKFSLYYQPRNAIVLVGVNYSOK 229
 DB 192 VVEDYLHAATAYQNTALXRTIFGPTENIKSISKDLDVYITTHYKGRIVYLAAGVSHDE 251
 QY 230 VPELSKRHEESLKNLDEKAIPTPYMKEPKODGARTAVHKGDVHLEWVALGYKVPKFK 289
 DB 252 LIDLAFHFHGDLSCTHKGELIPA--LPPCKFTGSEIR-VRODKMPLAHLAIAVEAVGMAR 308
 QY 290 DOVALALSRLLGEGKSSW-----LOSELVD---KKRLASQAFSHMQLODESVF--LFTA 340
 DB 309 DTICLWVANTLIGNMDRSPFGGNMSSKLAQLTCHGNLCHSFOSFTSYTDITGLMGLYV 368
 QY 341 GGNPNV--KAVALQKEIVALLKELKGEITQALDKIKINOKADFISSNESSDV 393
 DB 369 CESSIVADMHLHVQKEMWRLCTSVTESDVARAR-NILKTN---MLIQDGSPI 418

RESULT 8
 US-09-235-218-1
 Sequence 1, Application US/09235218
 Patent No. 6001629
 GENERAL INFORMATION:
 APPLICANT: Bandman, Olga
 APPLICANT: Shah, Putvi
 APPLICANT: Corley, Neil C.
 TITLE OF INVENTION: MITOCHONDRIAL PROCESSING PEPTIDASE
 TITLE OF INVENTION: SUBUNIT
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Inocyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/235, 218
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/895,521
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0343 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-855-0555
 TELEFAX: 415-845-4166
 TELEX:
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 489 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: BRAINOT14
 CLONE: 1593490
 US-09-235-218-1

Query Match 10.3%; Score 233; DB 3; Length 489;
 Best Local Similarity 23.9%; Pred. No. 1e-13;
 Matches 99; Conservative 80; Mismatches 196; Indels 40; Gaps 14;

QY 7 KRLLGSSVLLV--TLGASMH-----AQSYL--PKHESVTLKNGLOVSVPLE 50
 DB 16 RRLMGFSESLIRGAAGRSILYFGENRLRSTQATQVYLVNPEIRYTCLESGLRVASE--D 73
 QY 51 NKTGVLEVVDLYKVGSRNETMGKSGIAHMLEHLENFKSTNKLKAGEPDKYKRGVSNAS 110
 DB 74 SGISTCTVGLMIDAGSRYEENKNGTAHLEHNAFGTKKRSOLDLELEIENNGAHLNAY 133
 QY 111 TSPDITRYFKITSQANLMDKSLFPAETMGSNLKEDFLPERQVAAEERWRKTDNSPIGM 170
 DB 134 TSEQTVYVYAKAFSKDLPRAVELADIIONSTLGEAEIEREKVYLIREMOEVEFN--LQE 191
 QY 171 LYFRFEN-TAYVYHPYHWPPIGFMDIIONMTLDIRKFSLYYQPRNAIVLVGVNYSOK 229
 DB 192 VVEDYLHAATAYQNTALXRTIFGPTENIKSISKDLDVYITTHYKGRIVYLAAGVSHDE 251
 QY 230 VPELSKRHEESLKNLDEKAIPTPYMKEPKODGARTAVHKGDVHLEWVALGYKVPKFK 289
 DB 252 LIDLAFHFHGDLSCTHKGELIPA--LPPCKFTGSEIR-VRODKMPLAHLAIAVEAVGMAR 308
 QY 290 DOVALALSRLLGEGKSSW-----LOSELVD---KKRLASQAFSHMQLODESVF--LFTA 340
 DB 309 DTICLWVANTLIGNMDRSPFGGNMSSKLAQLTCHGNLCHSFOSFTSYTDITGLMGLYV 368
 QY 341 GGNPNV--KAVALQKEIVALLKELKGEITQALDKIKINOKADFISSNESSDV 393
 DB 369 CESSIVADMHLHVQKEMWRLCTSVTESDVARAR-NILKTN---MLIQDGSPI 418

RESULT 9
 US-09-134-001C-4950
 Sequence 4950, Application US/09134001C
 Patent No. 6380370
 GENERAL INFORMATION:
 APPLICANT: Lynn Doucette-Stamm et al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
 TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: GTC-007
 CURRENT APPLICATION NUMBER: US/09/134, 001C
 CURRENT FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 60/064,964
 PRIOR FILING DATE: 1997-11-08
 PRIOR APPLICATION NUMBER: US 60/055,779
 PRIOR FILING DATE: 1997-08-14
 NUMBER OF SEQ ID NOS: 5674
 SEQ ID NO 4950
 LENGTH: 424
 TYPE: PRT
 ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-4950

Query Match 9.3%; Score 210; DB 4; Length 424;
 Best Local Similarity 22.1%; Pred. No. 1.3e-11;

;; TITLE OF INVENTION: Proteases from Gram Positive Organisms
;; FILE REFERENCE: GC389-US
;; CURRENT APPLICATION NUMBER: US/09/554,999
;; CURRENT FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: PCT/US98/27040
;; PRIOR FILING DATE: 1998-12-17
;; PRIOR APPLICATION NUMBER: GB 9727471.6
;; PRIOR FILING DATE: 1997-12-30
;; NUMBER OF SEQ ID NOS: 3
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 2
;; LENGTH: 415
;; TYPE: PRT
;; ORGANISM: Bacillus subtilis
US-09-554-999-2

Query Match 10.4%; Score 236; DB 4; Length 415;

Best Local Similarity 24.6%; Pred. No. 4.1e-14;
Matches 103; Conservative 74; Mismatches 190; Indels 52; Gaps 15;

QY 40 NGLOVSVPLE--NKTGYEVDVLYKVSNETGKS-----GIAHMLEHLPKSTKN 90
DB 8 NGLOVYVLPKGFNKTYAVFTTKGSDINRFPYLGKNNMAYVPGIAHFLHKLPEKAD- 66
QY 91 LKAGEFDKIYKRFQGVSNASTSFDTIRYFKTSQANLDKSLFPAETGSLNKEDEFLP 150
DB 67 --GDVFQDFSKQASANAFTSFTRTAYLF--STSINVERNLEPLIDFQDPYFEKTYEK 122
QY 151 EROVVAERERRRTNSPIGMLYFRFPNTAYVYHRYHWPFGISFMDIQWTKDKIKFSL 210
DB 123 EKGITIGQINNYDN--PDMRLYGVNIENMYKHEVPRIDIACTASISHTIDLYECET 181
QY 211 YYQPKMAIVLVGVNDSQKPELSKHPESLKNLDEKAIPTPYKKEPQDQARTAVYHKD 270
DB 182 FYHPSNMILFYGVDPDAISQVRENGKKPYTDQPEIKKEVKE-----QNAVPEKE 235
QY 271 ---GVHLE---WVALGYKPAFHKDOVALDALSRLLGE---GKSSWLOSELVDKRLA 320
DB 236 KEIKMNVQGPCKLVGLKSNPKFKELKHELSNMLLEALFAKSS--AQYESLYEKGYI 294
QY 321 SQASHNMQLODESVLEFINGGNVKAALQKEIVALLETKEGETQALDKLKIYK 380
DB 295 DETSFPTAYGFGFAIGDTP--EPDQLAEDISSML--LRAGELTAKKIELARKKK 350
QY 381 -ADFISNLESSDVAGFADLYQNDIQGLTDYQROFLDKVSDLVRYANRY--EKDQOS 437
DB 351 IGTLLKALNSEFYIANQFTR-----AFIDMSLFDVYVLEQITLEDVON 395

RESULT 6
US-09-134-001C-4948
; Sequence 4948, Application US/09134001C
; Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4948
LENGTH: 445
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4948

Query Match 10.3%; Score 233.5; DB 4; Length 445;
Best Local Similarity 24.3%; Pred. No. 7.9e-14;

Matches 95; Conservative 62; Mismatches 187; Indels 47; Gaps 13;

QY 32 KHESTLKNGLGVSVPLENKTGYEVDVLY-----KVGSNETMGSGIAHM 79
DB 31 EHE--FDNGKLFTIP--KPGQKTYVYTTQPGSIDNHFKFGSQQFYKVPDVAHF 84
QY 80 LEHNFKSTKNLKAQFQKIVKRF--GCVNASTSFDTIRYFKTSQANLDKSLFPAE 136
DB 85 LEHLKFEK-----EDBEDLTAFAENQAQNAFTSPDRTYLF--SATSNIENSIKILN 136
QY 137 TMSLNLKEDEFLPEROVAEERRRRTDNSPIGMLYFRFPNTAYVYHRYHWPFGMDI 196
DB 137 MVEIPYFEETVYVNRKEGIIAEIKMYOE--QPGYKLMFETLIRAMSKPIRVADIGSVESI 195
QY 197 QNWTLKDKIKKHSILYQPKNAIVLVGVNDSQKPELSKHPESLKNLDEKAIPTPYKE 256
DB 196 YEITKDDLYLCYEFYHPSNNVLFVGVDSPOSTIKLVEKENORNTYQORIERAOIDE 255
QY 257 PQDQARTAVYHKDGVHLEWVALGYK-----VPAEKHKDOVALDALSRLLGKSSWL 309
DB 256 PREINQR-FVSEKKKILQSPRLMLQFKNEPDLDESATKFYQNRDLETFYELVGE--ETEFY 313
QY 310 QSELVDKRLASQAFSHNMQLODESVLEF--ACGNPVKAALQKEIVALLETKEGET 367
DB 314 QQLL--NKDLIDETFGQVLEPSYSFISITSATQOPDLQKOLIMDE---LRKYGNL 367
QY 368 TQAEIDKIKINQKADFTSNLESSDVAGLFA 398
DB 368 DQEAFLDKQFGEFISLSNPEYIANQYA 398

RESULT 7

US-08-895-521-1
; Sequence 1, Application US/08895521
; Patent No. 5869311

GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Shah, Puryi
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: MITOCHONDRIAL PROCESSING PEPTIDASE
TITLE OF INVENTION: SUBUNIT
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/895,521
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0343 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids
TYPE: amino acid

```
APPLICATION NUMBER: US/08/895,521
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0343 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 294589
US-08-895-521-3

Query Match
Best local Similarity 10.6%; Score 240; DB 2; Length 489;
Matches 97; Conservative 73; Mismatches 178; Indels 32; Gaps 13;

QY 30 LPRHESVTLKNGLOVSVPLENK-TGVIEVDLVYKGSRNEMTSGSIAMLEHLNFKST 88
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 55 VPETQVTCLENGLRVAS--ENSGISTCTVGLMIDAGSYRENEKNGTAHFLHMAFKGT 111
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 89 KNLKAGEFDKIVRFGVSNASTSPDITRYFITSQANLDSLELPAETMGSLNKEDEF 148
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 112 KRSQDLELEIMNGAHNAVTSRQTYVYAKAFSKDLPRAVEILLADIIONSTLGEAEI 171
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 149 LPRQVAAEERRWRDTPSGIMLYFRFEN-TAVYHPHYHTPIGEMDIONTLKIDIKKF 207
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 172 ERRRGVILREMOEVEN--LQEVVFDYLNATAYONTALGRTIIGPTENIKSISRDLVDY 229
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 208 HSLYQPKNAIVLVGDVNSQVFEISKHF-ESL--KNLDEKAIPTPYMKPKKODGART 264
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 230 ITTHYKGPRIVLAAGVCHNELLELAKHFHGDLSLCAHKGDVPLP-----PCKFTGSE 283
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 265 AVVHKGVHLEWALGYKPAFKHKDOVALDALSRLLGEGKSSW-----LQSELVD---K 316
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 284 IRYRDKMPRLAHVAIEAVGWTHTPTIRLMVANTLIGNDRSFGGGMNLSKLAOLTCH 343
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 317 KRLASQAFSHNMQLODESVF-LFIAGNPNV--KAEALQKEIYALLEKLKKEITQAEID 373
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 344 GNLCHSFQSFNTSYTDTGLMGLYMCEQATVADMLHAVOKEWMLCTAVASESEVARAK-N 402
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 374 KLIKOKADFIISLESSSDV 393
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 403 LKLTN---MLQLDGSSTPI 418
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
US-09-235-218-3
; Sequence 3, Application US/09235218
; Patent No. 6001629
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Puri
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: MITOCHONDRIAL PROCESSING PEPTIDASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Inocyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
```

```
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/235,218
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/895,521
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0343 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 294589
US-09-235-218-3

Query Match
Best local Similarity 10.6%; Score 240; DB 3; Length 489;
Matches 97; Conservative 73; Mismatches 178; Indels 32; Gaps 13;

QY 30 LPRHESVTLKNGLOVSVPLENK-TGVIEVDLVYKGSRNEMTSGSIAMLEHLNFKST 88
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 55 VPETQVTCLENGLRVAS--ENSGISTCTVGLMIDAGSYRENEKNGTAHFLHMAFKGT 111
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 89 KNLKAGEFDKIVRFGVSNASTSPDITRYFITSQANLDSLELPAETMGSLNKEDEF 148
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 112 KRSQDLELEIMNGAHNAVTSRQTYVYAKAFSKDLPRAVEILLADIIONSTLGEAEI 171
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 149 LPRQVAAEERRWRDTPSGIMLYFRFEN-TAVYHPHYHTPIGEMDIONTLKIDIKKF 207
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 172 ERRRGVILREMOEVEN--LQEVVFDYLNATAYONTALGRTIIGPTENIKSISRDLVDY 229
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 208 HSLYQPKNAIVLVGDVNSQVFEISKHF-ESL--KNLDEKAIPTPYMKPKKODGART 264
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 230 ITTHYKGPRIVLAAGVCHNELLELAKHFHGDLSLCAHKGDVPLP-----PCKFTGSE 283
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 265 AVVHKGVHLEWALGYKPAFKHKDOVALDALSRLLGEGKSSW-----LQSELVD---K 316
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 284 IRYRDKMPRLAHVAIEAVGWTHTPTIRLMVANTLIGNDRSFGGGMNLSKLAOLTCH 343
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 317 KRLASQAFSHNMQLODESVF-LFIAGNPNV--KAEALQKEIYALLEKLKKEITQAEID 373
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 344 GNLCHSFQSFNTSYTDTGLMGLYMCEQATVADMLHAVOKEWMLCTAVASESEVARAK-N 402
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 374 KLIKOKADFIISLESSSDV 393
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 403 LKLTN---MLQLDGSSTPI 418
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
US-09-554-999-2
; Sequence 2, Application US/09554999
; Patent No. 6465186
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
```


Best Local Similarity 23.3%; Pred. No. 9e-17;
Matches 102; Conservative 87; Mismatches 194; Indels 54; Gaps 14;

```
QY 26 AOSTLPKHESYTLKNGLOVSVPLENKTGVEVDLYLVKVSRRNETMKGSGIAHLEHNF 85
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 41 ALOEVPETQVSLDNGRLVASEOSSOPT--CTGVWIDVGRSPETKKNAGYPLEHLAF 98
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 86 KSTNKLAGEFDKIYKRRGGVSNASTSPDITRFRYFIKTQANLDSLELFAETMGSNLKE 145
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 99 KGTNRPGSALEKEYESGAILNASTREHTAYTIKALSKLPRAVELLGDIVONCSLED 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 146 DEFLEPQOVAEERRMRTDNPISGLMYFRFPNTAYVYHPYHTPIGEMDIONNTLADIK 205
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 159 SQIEKERDVIILREMQ--ENDASMRDVENVYLAHTAFQGTPLAQAVEGPEENVKLSRADLT 217
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 206 KFHSLYYQPKNAIYLVVDVNSQVFELESKHFSLEK--NLDEKALPTPYMEPPQDQART 264
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 218 EYLSTHYKAPRMVLAAGVGHQDLDLAQKHGLGIPWYAEADVPT--LTPCRFTGSE- 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 265 AVVHK--DGVHLEWVALGYKVPAPFKHKDOVALDALSLRLGE-----GKSSMLQSELY 314
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 275 -IRRDALPPRAHYAIAVEGGMASPDVALQVANAIGHDCITYGGVHLSPLASGAV 333
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 315 DKRLAS--QAFS-----HNMQLODESVFLFTAGGNPNYKABALQKEIYAL 358
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 334 ANKLQSFQTFPSICVAETGLGAHFVCDRMKID--MMFV-----LOGQWRML 379
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 359 LEKLKGEITQAEIDKLKINOKADPISNLESSDYA--GLFADYLVQNDIOGLTYQRF 417
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 380 CTSATESEVARGK---NILRNA--LVSHLDGTPVCEDIGRSLTYGRRIPLAEMESRIA 434
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 418 DLKVSIDLVRANEXEKD 434
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 435 EVDASVAREICSKITYD 451
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 2
US-09-235-218-4
Sequence 4, Application US/09235218
Patent No. 6001629

GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: MITOCHONDRIAL PROCESSING PEPTIDASE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/235, 218
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/895,521
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PR-0343 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166

TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:

LENGTH: 480 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1082896
US-09-235-218-4

Query Match 11.7%; Score 265; DB 3; Length 480;
Best Local Similarity 23.3%; Pred. No. 9e-17;
Matches 102; Conservative 87; Mismatches 194; Indels 54; Gaps 14;

```
QY 26 AOSTLPKHESYTLKNGLOVSVPLENKTGVEVDLYLVKVSRRNETMKGSGIAHLEHNF 85
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 41 ALOEVPETQVSLDNGRLVASEOSSOPT--CTGVWIDVGRSPETKKNAGYPLEHLAF 98
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 86 KSTNKLAGEFDKIYKRRGGVSNASTSPDITRFRYFIKTQANLDSLELFAETMGSNLKE 145
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 99 KGTNRPGSALEKEYESGAILNASTREHTAYTIKALSKLPRAVELLGDIVONCSLED 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 146 DEFLEPQOVAEERRMRTDNPISGLMYFRFPNTAYVYHPYHTPIGEMDIONNTLADIK 205
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 159 SQIEKERDVIILREMQ--ENDASMRDVENVYLAHTAFQGTPLAQAVEGPEENVKLSRADLT 217
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 206 KFHSLYYQPKNAIYLVVDVNSQVFELESKHFSLEK--NLDEKALPTPYMEPPQDQART 264
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 218 EYLSTHYKAPRMVLAAGVGHQDLDLAQKHGLGIPWYAEADVPT--LTPCRFTGSE- 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 265 AVVHK--DGVHLEWVALGYKVPAPFKHKDOVALDALSLRLGE-----GKSSMLQSELY 314
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 275 -IRRDALPPRAHYAIAVEGGMASPDVALQVANAIGHDCITYGGVHLSPLASGAV 333
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 315 DKRLAS--QAFS-----HNMQLODESVFLFTAGGNPNYKABALQKEIYAL 358
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 334 ANKLQSFQTFPSICVAETGLGAHFVCDRMKID--MMFV-----LOGQWRML 379
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 359 LEKLKGEITQAEIDKLKINOKADPISNLESSDYA--GLFADYLVQNDIOGLTYQRF 417
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 380 CTSATESEVARGK---NILRNA--LVSHLDGTPVCEDIGRSLTYGRRIPLAEMESRIA 434
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 418 DLKVSIDLVRANEXEKD 434
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 435 EVDASVAREICSKITYD 451
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 3
US-08-895-521-3
Sequence 3, Application US/08895521
Patent No. 5869311

GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: MITOCHONDRIAL PROCESSING PEPTIDASE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 11:56:49 ; Search time 18 seconds
(without alignments)
725.765 Million cell updates/sec

Title: US-09-881-752a-212
Perfect score: 2270
Sequence: 1 MKHFSVKRLGLSSVLVTL.....VRANEYFKDOSTVFLKP 444

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/pdata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/pdata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/pdata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/pdata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/pdata/1/1aa/6C.COMB.pep:*
6: /cgn2_6/pdata/1/1aa/6D.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	265	11.7	480	2	US-08-895-521-4
2	265	11.7	480	3	US-09-235-218-4
3	240	10.6	489	2	US-08-895-521-3
4	240	10.6	489	3	US-09-235-218-3
5	236	10.4	415	4	US-09-554-999-2
6	233.5	10.3	445	4	US-09-134-001C-4948
7	233	10.3	489	2	US-08-895-521-1
8	233	10.3	489	3	US-09-235-218-1
9	210	9.3	424	4	US-09-134-001C-4950
10	188	8.3	1259	4	US-09-187-049-13
11	185.5	8.2	1019	4	US-09-434-066-23
12	151	6.7	124	3	US-09-187-049-8
13	145.5	6.4	136	3	US-09-187-049-6
14	141.5	6.2	125	3	US-09-187-049-7
15	126	5.6	129	3	US-09-187-049-5
16	115.5	5.1	649	4	US-09-134-001C-3891
17	109.5	4.8	1049	2	US-08-817-090B-2
18	109.5	4.8	1050	2	US-08-817-090B-4
19	109.5	4.8	1101	2	US-08-916-917-14
20	109.5	4.8	1101	3	US-09-225-170-14
21	108.5	4.7	129	3	US-09-187-049-3
22	107.5	4.7	1073	4	US-09-541-782-6
23	107.5	4.7	1073	4	US-09-723-820-6
24	107	4.7	755	4	US-09-097-767A-38
25	106.5	4.7	334	4	US-09-071-035-106
26	105	4.6	1248	2	US-08-348-353-17
27	105	4.6	1248	2	US-08-465-965-17

28	105	4.6	1248	3	US-08-465-966-17
29	104.5	4.6	1095	4	US-09-206-942-45
30	104.5	4.6	1101	4	US-09-206-942-43
31	103	4.5	922	2	US-08-464-402-2
32	103	4.5	922	4	US-09-054-775C-2
33	102.5	4.5	297	3	US-09-187-049-10
34	102.5	4.5	983	4	US-09-412-554A-2
35	102.5	4.5	1102	2	US-08-916-917-4
36	102.5	4.5	1102	2	US-08-972-631-4
37	102.5	4.5	1102	2	US-08-972-629-4
38	102.5	4.5	1102	2	US-08-972-630-4
39	102.5	4.5	1102	2	US-08-672-211-4
40	102.5	4.5	1102	3	US-09-225-170-4
41	101	4.4	165	4	US-09-554-999-3
42	99.5	4.4	268	3	US-09-187-049-9
43	99.5	4.4	655	4	US-09-632-538C-36
44	99.5	4.4	905	4	US-09-134-001C-3782
45	99.5	4.4	1164	4	US-08-923-992A-10

ALIGNMENTS

RESULT 1
US-08-895-521-4
; Sequence 4, Application US/08895521
; Patent No. 5869311
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: MITOCHONDRIAL PROCESSING PEPTIDASE
; TITLE OF INVENTION: SUBUNIT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/895,521
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0343 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1082896
; US-08-895-521-4
Query Match 11.7%; Score 265; DB 2; Length 480;

Downloaded from

Db 287 MLIGWTSKIYHTLYEDGK-----IDSEFVDVVEIHNEQF---VLISLDTPEPIAM 334
QY 348 AEALQREIVALLEKLEKKGEITQAEIDKIKINQKADFTSNLESSDVAGLFADYLVQNDIQ 407
Db 335 SNYIRQKLIAT--KISK-EFTNEHLNLKKEMWGDFTIOSLDSIEHLTHQFSLYLSDSDE 391
QY 408 GLTDYQROFLDLKVSDDLVRVANEYFK--DTQSTYVFLK 443
Db 392 TYFDIPKIIERLTLKDVVTIGKAFPEKADASDFTVFPK 429

Search completed: November 18, 2002, 11:57:35
Job time : 43 secs

PT Novel nucleic acids encoding 49 human secreted proteins useful for
 PR treating cancers, hyperproliferative disorders, inflammatory disorders,
 PR neurological disorders and cardiovascular disorders -
 PS Disclosure: Page 385-387; 389pp; English.

CC The invention relates to the isolation of genes AAC59108-C59156 encoding
 CC the human secreted proteins AB28012-B28060. This sequence represents a
 CC peptide fragment homologous to the protein encoded by the gene given
 CC in the descriptor line. The sequence is a search result from a BLASTX
 CC homology search. The genes and proteins are useful for preventing,
 CC ameliorating or treating medical conditions, e.g. by protein or gene
 CC therapy. The genes are isolated from a range of human tissues disclosed
 CC in the specification. The nucleic acids, proteins, antibodies and
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
 CC (a) cancer, e.g. breast and ovarian cancer, and other cancers of the
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
 CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; (c) cardiovascular disorders such
 CC as myocardial ischaemia; (d) wound healing; (e) neurological diseases
 CC e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such
 CC as viral, bacterial, fungal and parasitic infections.

CC Sequence 462 AA:

Query Match 10.4%; Score 237; DB 21; Length 462;
 Best Local Similarity 24.1%; Pred. No. 3.2e-11;
 Matches 100; Conservative 80; Mismatches 195; Indels 40; Gaps 14;

QY 7 KRLLGLSVLV--TLGASMH-----AQSYL--PKHESVTLKNGLOVSVPLE 50
 DB 9 RLIMFSESLLRGAAGRLVFGENLRSTOATQVVLNVPFRVYLCESGLRVASE--D 66
 QY 51 NKTGVLEVDLVKVSRRNEMGSGIAHMLEHLNFKSTKNLAGFEKIVKRGVSNAS 110
 DB 67 SGLSTCTVGLMIDAGSRYENKKNNGTAHLEHMAFGTKRSQDLDELEINMGALNAY 126
 QY 111 TSPFDITRYIKTSQANLDSLFLFAETMSLMLKEDEFLPEQOVAEERKRIQNSPIGM 170
 DB 127 TSREQTVVYAKAFSKDLPAVEILADIIONSTLGEAIEIRRGVILREMOVEFTN--LDE 184
 QY 171 LYERFEN--TAYVYHPHMTPIGFMDIQMTLKDIFKFSLYOPKNAIVLVGDVNSOK 229
 DB 185 VVEDYIHAATAYQNTALGRITLGTPTENIKSIRKDLVDYITTHYKGRIVLAAGVSHDE 244
 QY 230 VELSKEHESLKNIDEKAIPTPYMKPEKQDGAFTAVHKGDVHLEWALGYKVPFKHK 289
 DB 245 LLDLAKFHEGDSLCYHKGEP--LPCKFTGSEIR--VRDDKMPILNHLAIAVAVGMHP 301
 QY 290 DOVALDALSRILGEGKSSM-----LOSELVD---KKRLASQAFSHNMQLODESVF--LFTA 340
 DB 302 DTICLMAVNTLLGNMDSRFGGGMNLSKLAQLTCHGNLCHSFQSFMTSYTDIGLMLYAV 361
 QY 341 GGNPNV--KAALQKEIYVALLKELKGETQALDKLINOKADFSNNESSSDV 393
 DB 362 CESSIVADMLHVQKEMKRLCTSVIESEVAVAR--NLTKTN-----MLLOLDGSPFI 411

RESULT 15
 ABP26992
 ID ABP26992 standard; Protein; 429 AA.

XX ABP26992;
 AC 02-JUL-2002 (first entry)
 XX Streptococcus polypeptide SEQ ID NO 3160.
 DE Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KM group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.

XX Streptococcus pyogenes.
 OS WO200234771-A2.
 XX 02-MAY-2002.
 PD 29-OCT-2001; 2001WO-GB04789.
 PF 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.

PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX Telford J, Maignant V, Margarit Ros YI, Grandi G, Fraser C;
 PI Telford J;
 DR WPI; 2002-352536/38.
 DR N-PSDB; ABN67623.

PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 PS Claim 1; Page 3469; 4525pp; English.

CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.

CC Sequence 429 AA:

Query Match 10.4%; Score 236.5; DB 23; Length 429;
 Best Local Similarity 23.1%; Pred. No. 3.2e-11;
 Matches 106; Conservative 72; Mismatches 183; Indels 97; Gaps 18;

QY 36 VTLKNGLOVSVPLENKTVIEVDLVKVG-----SRNETGKSGIAHMLEH 82
 DB 19 VLENGLVYFYI--KKIFLEKTAMLTVGFSLDNKLTVDDESRI--APAGIAHFLSH 72
 QY 83 LMFKSTKNLAKGEPDQIVKRGVSNASPSDITRREITSQANLDSKLELAETMGSLN 142
 DB 73 KLFEDF--SGDDISLKFQDLGAETNAFTFMQTSFFSSTA--SKFOENLELLOYVLSAN 128
 QY 143 LKEDFLPERQVVAEERRRRTNSPI---GMLYERFFNTAVYVHPMTPIGFMDION 198
 DB 129 ITDESVSREKTIIGQELIDYQDDADVRAVSGILQNFPTISLAN-----DIAGSKASIOK 183
 QY 199 WTLKDIKFFSHLYOPKNAIVLVGDVNSQKVELSKHFESLKNIDEKAIPT-----P 252
 DB 184 ITKILLETHTHTYFQPTNMSLFTVGDIDIDEFLAIQRFQTLSPDRKRYVVDPLHYXP 243
 QY 253 YKKEKQD-----GARTPAVYHKGDVHLEWALGYKVPFKKKDOVALDA--LS 298
 DB 244 VIKSSVDMVDVTTAKLVGFRGLVLTQHS-----LITFR-----TALKLFVS 286
 QY 299 RLIG-----EGKSSWLQSELVDRKKRLASQAFSHNMQLODESVFLETAAGNPVK 347

DB 192 HKDFYQLSNNMLVVGQFSPQETITTYLQKNSHPTSYSONIDRDISL-----EP----- 240
QY 265 AAVHKRDVHL-----EWVALGK-----VPAFKHKQVALDA-LSRLDGEKGS---SMLQS 311
DB 241 -VINKNSCHMTVTPKPAIKGKSNMHMHSYLYKEKIGLOFFMMLGWTINQDMYES 299
QY 312 ELVDKRLASQAFSHNQLODESVFLFIAGNPVKAALQKEIVALLEKKEITQAE 371
DB 300 GQID-----DSFDIEIEVHDFECVILISDTEPIAFSTQLRL-LKNALQSSDLNESH 352
QY 372 LDKLKINQKADFISSLSSDVACLFADYLVQNDIQGLTYQRFLO-----LKVS DL 424
DB 353 LKNVKKRELGGDFLRSLSIENLAMOFTYLY-----DCKIMYLDLPSTIVEELDEDV 404
QY 425 VRANEXE--KDTQSTVFLK 443
DB 405 ITICKDFLDNADTSDVFIIPK 425
RESULT 13
AAB28098
ID AAB28098 standard; protein; 460 AA.
AC AAB28098;
02-FEB-2001 (first entry)
Human secreted protein BLAST search protein SEQ ID NO: 146.
DE
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnereary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX
OS Homo sapiens.
XX
PN WO200055177-A2.
XX
PD 21-SEP-2000.
XX
PE 09-MAR-2000; 2000WO-US06058.
XX
PR 12-MAR-1999; 99US-0124145.
PR 03-DEC-1999; 99US-0168654.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-638177/61.
XX
XX Novel nucleic acids encoding 49 human secreted proteins useful for
XX treating cancers, hyperproliferative disorders, inflammatory disorders,
XX neurological disorders and cardiovascular disorders
XX
XX Disclosure: Page 387-388; 389pp; English.
XX
XX The invention relates to the isolation of genes AAC29108-C59156 encoding
XX the human secreted proteins AAB28012-B28060. This sequence represents a
XX fragment of the protein encoded by the gene given in the descriptor
XX line. The sequence is used as a query sequence for doing BLASTX searches
XX to determine homologous sequence to the protein. The genes and proteins
XX are useful for preventing, ameliorating or treating medical conditions,
XX e.g. by protein or gene therapy. The genes are isolated from a range of
XX human tissues disclosed in the specification. The nucleic acids,
XX proteins, antibodies and (ant)agonists are useful in the diagnosis,
XX treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer,
XX and other cancers of the adrenal gland, bone, bone marrow, breast,
XX gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders
XX e.g. Addison's disease, allergies, autoimmune haemolytic anaemia,
XX autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple
XX sclerosis, rheumatoid arthritis and ulcerative colitis; (c)

CC cardiovascular disorders such as myocardial ischaemias; (d) wound
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and
CC (f) infectious diseases such as viral, bacterial, fungal and parasitic
CC infections.
XX
SQ Sequence 460 AA;
Query Match 10.4%; Score 237; DB 21; Length 460;
Best Local Similarity 24.1%; Pred. No. 3.2e-11;
Matches 100; Conservative 80; Mismatches 195; Indels 40; Gaps 14;
QY 7 KRLIGLSSVLLV--TLGASH-----AQSIL--PKHBSVTLKGLQVSVLE 50
DB 7 RRLWGFESILLIRGAAGRSILYFGENRLRSTQAAQVVLNDETHVTCLESLRVASE--D 64
QY 51 NKTGIEVDVLYKVGSRNEMPKSGIAHMLHMLFKSTKNLKAEPDKYKRGCVSNAS 110
DB 65 SGLSTCTVGLWIDAGSRNENKNGTAFLEHMAFKGTRKRSQDLLELEENMGAHNLAY 124
QY 111 TSPDITRYFIKTSQANDKSLLEPAETMGSLNKEDEFLPERQVAAERMRRTNSPIGM 170
DB 125 TSREGIYVAKAFSKDLPRAVEILADILQNSTLEAEIERRGVILKEMQVEFN--LQE 182
QY 171 LYRFEFN-TAVVHPHYMTPIGFMDDIQMTLKDIKKPHSLYQPKNAIYLVGDVNSOK 229
DB 183 VFEDYLHATAVQNTALGRTILGPTENIKSISRKDLVDYITTHYKGPRIVLAAGVSHDE 242
QY 230 VFELSKHFEELKMLDEKAITPTPKPKKODGATAVVHNDGVILLEVALGKYPAPKKH 289
DB 243 LLDIAKPHGDSLCTHNGEIPA--LPCKFTGSEIR-VRODKMPLAHIAVEAVGNMAH 299
QY 290 DQVALDLASRLRLEGKSSW-----LQSELYD---KKRLASQAFSHNQLODESVF-LFIA 340
DB 300 DTCLMVAANTLIGWMDSPFGGMLSSKIALQPLCHGMLCHSFQSFNRYDTGLMGLXNV 359
QY 341 GGNPNV--KAAALQKEIVALLEKKEITQAEIDKLIKINQKADFISSLSSDV 393
DB 360 CESSTVADMLHVQKEMWRILCTSVTESEVARAR-NLKTN---MLLDGSGTPI 409
RESULT 14
AAB28097
ID AAB28097 standard; protein; 462 AA.
AC AAB28097;
02-FEB-2001 (first entry)
Human secreted protein BLAST search protein SEQ ID NO: 145.
DE
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnereary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX
XX Homo sapiens.
XX
PN WO200055177-A2.
XX
PD 21-SEP-2000.
XX
PE 09-MAR-2000; 2000WO-US06058.
XX
PR 12-MAR-1999; 99US-0124145.
PR 03-DEC-1999; 99US-0168654.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-638177/61.
XX

PR 31-DEC-1997: 97AU-0001182.
PR 30-JAN-1998: 98AU-0001546.
PR 10-MAR-1998: 98AU-0002264.
PR 09-APR-1998: 98AU-0002911.
PR 23-APR-1998: 98AU-0003128.
PR 05-MAY-1998: 98AU-0003338.
PR 22-MAY-1998: 98AU-0003654.
PR 29-JUL-1998: 98AU-0004917.
XX
XX (CSLC-) CSL LTD.
XX
XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
XX
XX WPI, 1999-385613/32.
XX N-PSDB; AAX91553.
XX
XX Antigenic Porphyromonas gingivalis peptides for preventing
PT gingivitis
XX
XX Claim 1, Page 293-295; 588pp; English.
XX
XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (Pg) polypeptide sequences given in AAX94318 to
CC AAX94583. AAX91802 to AAX91989 represent PCR primers used in the
CC isolation of the Pg polypeptides. The Pg polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The Pg polypeptides can be
CC used as vaccines especially against Porphyromonas gingivalis. Probes can
CC be used to detect Porphyromonas gingivalis in standard hybridisation
CC assays. Porphyromonas gingivalis is involved in periodontal disease
CC especially gingivitis.
XX
XX Sequence 945 AA:
SQ
Query Match 11.6%; Score 262.5; DB 20; Length 945;
Best Local Similarity 26.8%; Pred. No. 6.7e-13;
Matches 111; Conservative 56; Mismatches 184; Indels 63; Gaps 14;
QY 6 VKRLGISLVLYTLGASMAQSY--LPKHESV---TLKGLGVSVPLNKNKGIVIVDV 60
DB 7 LKRFLLSGLLVGFIPKLSAQQAQPLPDPRAVRGKLDNGLTYYTRHNENPDRADFETI 66
QY 61 LYVGSNFTMGSGIAHMLEHLNFKSTKNLKA---GEFDKIYKRGVSNASTSPDIT 116
DB 67 AOKVGSILEDSSQGLAHFLEHMAFNKTRPKPKNLINTYETIGVRGQMLNASTGDKT 126
QY 117 RYFL---KTSQANLDSLELFAETWGSNLKEDFLPERQVVAEERMTDNSPIGLMY 172
DB 127 EYTIMDVPTTRGIDISCLITLDWSNNITLDGHEIDEERGVIQOEWRARRDAN----- 180
QY 173 FREFNT---AAYYHRY-HHTPIGFMDIDIONWTLKDKRHSLYQPKNAIVLVGVNS 227
DB 181 LRMFEALAKAMPGNKIAEHRPIGLMDVNLNFHDELRLNKKWYRPDLQGLVIVGID- 239
QY 228 OKVFELSKHFESLKNDEKAIPTPYKKEPKODGARTAVVHKDGNHLEWALGKYVAFK 287
DB 240 --VDYENKIKLEKQVPAVNPAPERITYTVEVDENDEPIVALDADEFTTQLST---SFK 294
QY 288 -----HKDVALDALSRLLGEGKSSWLSQSELVDKR--LASQAFSH 326
DB 295 SDPTPQEVGRSIFGLVEDYMKQVITTAVERL-----SEITHKPNAPFLSAGAFS 345
QY 327 NMQ--IQDESVFLFIAGGNPNVKAELQKEIVALLEKLGKGETTQALDKLTKIN 378
DB 346 NFMVITQTKDAFNFAVATVREG-EAEKAMNALVAEIESLROFGITKGEYDRARTN 398
RESULT 12
ABP26991
ID ABP26991 standard; Protein; 427 AA.
XX
XX AC ABP26991;
XX

DT 02-JUL-2002 (first entry)
XX
XX Streptococcus polypeptide SEQ ID NO 3158.
DE
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX group A streptococcus; Streptococcus pyogenes; antibacterial;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
KM antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
XX Streptococcus agalactiae.
OS
XX WO200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB04789.
XX
XX 27-OCT-2000; 2000GB-0026333.
XX 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
XX (CHR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
XX Telford J, Maignan V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettelin H;
XX
XX WPI, 2002-352536/38.
DR N-PSDB; ABN67622.
XX
XX Claim 1, Page 3469; 4525pp; English.
PS
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
XX Sequence 427 AA:
SQ
Query Match 11.5%; Score 260; DB 23; Length 427;
Best Local Similarity 23.4%; Pred. No. 3.5e-13;
Matches 103; Conservative 91; Mismatches 181; Indels 66; Gaps 20;
QY 36 VTKNGIQVSVSP---LENKTVIEVDV--LYVGSNFTMG--KSGIAHMLEHLNFKST 88
DB 18 LTLESGLNLYLIRPKSKETVGLTANFSGLSHKKYRNGCVEHYPAIGAHFELHKLFEID 77
QY 89 KNLKAG-EFDKIYKRGVSNASTSPDITRYFKTSQANLDSLELFAETWGSNLKED 147
DB 78 KGODAATQFTK---YGAESNAFTFDKTSFYSTI-SHTNCLDILLDFVLTPNTEES 132
QY 148 FLPERQVVAEERMTDNSPIGLMYRFRFTAVVHPYHHTPIGFMDIDIONWTLKDKR 207
DB 133 ITREKDIIOKEIEMYQDD-PEYRLYQGVLSNLKPNSEPLAFDIAGDYQSISQITLTDLEN 191
QY 208 HSLYQPKNAIVLVGVNSQKVFELSK--HFESL-KNLDEKAIPTPYKKEPKODGART 264

Query Match 12.5%; Score 283.5; DB 22; Length 438;
Best Local Similarity 27.5%; Pred. No. 4e-15;
Matches 111; Conservative 66; Mismatches 201; Indels 25; Gaps 9;

37 TLKNGLOVVS--VPLEKKTGVIEVDVLYKVGSRNETMCKGSIAMLEHLNFKSTKNLAKG 94
19 TLGEGLRVTEELFPAVNSA---SYGVWVGVGSRDGATVGAHAFLEHLFKSTPTRSAV 75
95 EPDKIVRFGVSNASTSFDTIRYFIKTSOANLDKSLLEFAETGSIINLKEDEFLPRROY 154
76 DIAQMDAVGELNAFLAKKHTCTYAHVLSGSDPLAVDLVADVVLNRCADADVEVERDY 135
155 VAERRRRTDNPISGIMLYFRFNTAYVYHPYHWPICFMODIQWTLKDKIKFHSLLYQP 214
136 VLEETIAR--DDPEDALADMEFLALFGDHPGVRPISGASVSQSVYTRAQLOSFILRRYTP 194
215 KNIIVLVGVDSQKVELSKHESLKNDEKAIPTPYKKEPKQDA--PRAVYHKKGVH 273
195 ERVYVAAAGVNDHDLVALVREHFGSLVNGRRPV-APRKGTGVNCSPLRTLVSRDAEQ 253
274 LEWVALGYKVPAFKHKDOVALDALSRLLGKSSWLQSELVQDKKRLASQAFSHNQLODE 333
254 TH-VSLGIRTPRGWEHRLMSLVLTALGGLSRLFOEVRETGLAVSYS----- 304
334 SVLEFLIAGNPYKAKALQKEI-----VALLERLKGELTQALDKLKNORADEISN 386
305 ALDIEFADSGALSYAACLPREFADVMEKVTADVLESVALRDGITTEACGIANGSLNGLVLG 364
387 LE-SSSDVAGLFDVLYQNDIQGLTDYQROFLDKVSDLYRVA 428
365 LEQSSSRMSRLGSELMYNGKHSIEHTLRQIEQVTEEVNAYA 407

Db

RESULT 10
AAV34459
ID AAV34459 standard; Protein; 941 AA.
XX
AC AAV34459;
XX
DT 25-AUG-1999 (first entry)
XX
DE Porphyromonas gingivalis protein Pg115.
XX
KW Porphyromonas gingivalis; Pg; periodontal disease; gingivitis;
XX
KM vaccine; antigenic.
XX
OS Porphyromonas gingivalis.
XX
PN W0929870-A1.
XX
PD 17-JUN-1999.
XX
10-DEC-1998; 98W0-AU01023.
XX
04-AUG-1998; 98AU-0005028.
XX
10-DEC-1997; 97AU-0000839.
XX
31-DEC-1997; 97AU-0001182.
XX
30-JAN-1998; 98AU-0001546.
XX
10-MAR-1998; 98AU-0002264.
XX
09-APR-1998; 98AU-0002911.
XX
23-APR-1998; 98AU-0003128.
XX
05-MAY-1998; 98AU-0003338.
XX
22-MAY-1998; 98AU-0003654.
XX
29-JUL-1998; 98AU-0004917.
XX
PA (CSLC-) CSL LTD.
XX
PI Agius CT, Barr IG, Hocking DM, Margetis MB, Patterson MA;
XX
PI Ross BC, Rothel LJ, Webb EA;
XX
DR WPI: 1999-385613/32.
XX
DR N-PSDB; AAV91677.
XX

PT Antigenic Porphyromonas gingivalis peptides for preventing
PT gingivitis
XX
PS Claim 1; Page 436-437; 588bp; English.
XX
XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (Pg) polypeptide sequences given in AAV34318 to
CC AAV34583. AAX91802 to AAX91989 represent PCR primers used in the
CC isolation of the pg polypeptides. The pg polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The pg polypeptides can be
CC used as vaccines especially against Porphyromonas gingivalis. Probes can
CC be used to detect Porphyromonas gingivalis in standard hybridisation
CC assays. Porphyromonas gingivalis is involved in periodontal disease
CC especially gingivitis.
XX
SQ Sequence 941 AA;

Query Match 11.6%; Score 262.5; DB 20; Length 941;
Best Local Similarity 26.8%; Pred. No. 6.7e-13;
Matches 111; Conservative 56; Mismatches 184; Indels 63; Gaps 14;

6 VKRLGSSVLVLYLGASMAHQS--LPKHESV---TLKNGLOVVSVPLEKKTGVIEVDV 60
3 LKRLSLGILLVGEIIPKLSAQOAPLPTDPVAVGVKIDGLTYFIRHNENPKDRADFFI 62
61 LYKVSNETMCKGSIAMLEHLNFKSTKNLKA----GERDKIVRFGVSNASTSFDT 116
63 AQKVGSLLEDSOSGLAHFLEHMAFNCTKNFPGKNLINYLETIGVRGQMLNSTGFDKT 122
117 RYFI---KTSQANLDKSLLEFAETGSLNLEDEFLPERQVVAERRRRTDNPISGIMLY 172
123 EYTIMDVPTTRQGITDSCILIHDMWSNNITLDGHEIDEERGVIQIEWRARRDAN----- 176
173 FRFNT---AAYVHPT-HNTPICFMODIQWTLKDKIKFHSLLYQKNAIVLYGVDSN 227
177 LRMEFLAIKAMPENKKAERMPICIMDVNLNFKDELRNYYKKWYRBDLGLVIVGDI- 235
228 QKVELSKKHESLKNDEKAIPTPYKKEPKQDAPRTAVYHKKDGVHLEWVALGYKVPAFK 287
236 --VDYVENKIKELFKDVPAPVNPAPERIYTYVEDDEPIVALDAEATTTQOLST---SFK 290
288 -----HKDOVALDALSRLLGKSSWLQSELVQDKK--LASQAFSH 326
291 SDPTPQEVRSIGLVDMYKQVITTAVERL-----SETTHKPNMPISAGAFS 341
327 NMQ--LQDESVEFLIAGNPYKAKALQKEIVALLEKLGELTQALDKLKN 378
342 NFMYITQTKDAFNFVATVREG-EAEKAMNALVAIEISLRQGITKGYDRARTN 394

Db

RESULT 11
AAV34335
ID AAV34335 standard; Protein; 945 AA.
XX
AC AAV34335;
XX
DT 25-AUG-1999 (first entry)
XX
DE Porphyromonas gingivalis protein Pg115.
XX
KW Porphyromonas gingivalis; Pg; periodontal disease; gingivitis;
XX
KM vaccine; antigenic.
XX
OS Porphyromonas gingivalis.
XX
PN W0929870-A1.
XX
PD 17-JUN-1999.
XX
10-DEC-1998; 98W0-AU01023.
XX
04-AUG-1998; 98AU-0005028.
XX
10-DEC-1997; 97AU-0000839.
XX

```

PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154079.
PR 20-SEP-1999; 99US-0154739.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 12.8%; Score 290; DB 21; Length 503;

Best Local Similarity 24.3%; Pred. No. 1,4e-15; Indels 44; Gaps 13;

Matches 104; Conservative 84; Mismatches 196;

```

QY 31 PKHSEYLNKNOVSVLP-LEKRTGVIEVDLYKVGSRNETMKSGIAHMLNFKSTK 89
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 95 PERRVITLPGCLVATNESNSAKTAT--VGWVIDAGSRFESDSTNGAHLEHMIKFGTD 152
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 90 NKAAGEBDKIVKRFGGVSNASTSFDTIRFIKTSQANLDSLELPAETMGSMLKDEFL 149
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 153 RRTVRLALEEIEDIGHLNAYTSREQTYVAKYLDNSVNOALDVLADIIONSKFEQRIN 212
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 150 PERQVAAEERR--WRDINSPIGMLFRFENTAYVYHPYWPFGIMDDIQNTLMDIKK 206
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 213 RERDVLIREQVEVGQTDDEVLDHLH---ATAFYTPGLGRTLLGPAQNVKSITREDLON 268
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 207 FHSLEYOPKNAIVLVGDVNSOKVFLSKHFSLNIDEKALPTPYMKREKODGARTAV 266
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

DB 269 YIKTHYASRMVIAAGAVKHEEVEVQVKLFETKLS-DPTTTSOLVANEPASFTGSEVR 327
QY 267 VHKDGVHLEWVALGYKVPAPKHKDOVALDALSRLLGE-----GKSNLOSSELVDK---KR 318
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 328 MIDDDELPLAOFVAVFEGASWTDPPDSVALVMQTMGSNNKNGGCKHVSDDLQRYAINE 387
QY 319 LASQAFSHNMQLODESVEFLFIAGNPNVKAERALQEIYALLERLKK--GEITQAEIDKIK 376
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 388 IAESIMAFNTNKKDGLF-----GYAVAKADCLDLDLSYIMVEYKRLAVRSDADYTRAR 443
QY 377 INOKADEFISNELSSSDVAGLFADYLVQNDI-OGLDYQRF-----LD-LKYSDLVR 426
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 444 NOKSSILLHMDGTSPBIA-----EDICRQLLYGRIRPTAELFARIDAVDASTVKR 494
QY 427 VANEPFKD 434
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 495 VANKYLYD 502

```

```

RESULT 9
AAG81200
ID AAG81200 standard. Protein; 438 AA.
XX
AC AAG81200;
XX
DT 04-SEP-2001 (first entry)
XX
DE Mycobacterium tuberculosis potential drug target protein SEQ ID 251.
XX
KW Drug target; growth; organism viability; characterisation.
XX
OS Mycobacterium tuberculosis.
XX
PN W0200135317-A1.
XX
PD 17-MAY-2001.
XX
PF 13-NOV-2000; 2000MO-US31152.
XX
PR 12-NOV-1999; 99US-0165086.
PR 12-NOV-1999; 99US-0165124.
PR 01-FEB-2000; 2000US-0179531.
XX
PA (REGC ) UNIV CALIFORNIA.
XX
PI Eisenberg D, Rotstein SH, Marcotte EM;
XX
DR N-PSDB; AAH52051.
XX
PT WPI; 2001-329193/34.
XX
PT Identifying nucleotide or polypeptide sequence for use as drug target,
PT involves providing algorithm that analyzes a functional relationship
PT between nucleotide or polypeptide sequences, and comparing the
PT sequences
XX
PS Disclosure: Page 181; 207pp; English.
XX
CC This invention relates to a method for identifying a nucleotide or
CC polypeptide sequence that may be a drug target, or essential for growth
CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
CC tuberculosis proteins which are potential drug targets. The DNA and
CC protein sequences are used to illustrate the method of the invention. The
CC method involves providing an unknown nucleotide or polypeptide sequences,
CC and comparing it to a number of sequences along with at least one
CC algorithm capable of analysing a functional relationship between
CC nucleotide and polypeptide sequences. The method is useful for
CC characterising the function of nucleic acids and polypeptides that may be
CC useful as a target for a drug or essential for the growth or viability of
CC an organism.
XX
SQ Sequence 438 AA;

```


ID AAG35643 standard; Protein; 503 AA.
XX
AC AAG35643;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 43573.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 9905-0121825.
PR 05-MAR-1999; 9905-0123180.
PR 09-MAR-1999; 9905-0123548.
PR 23-MAR-1999; 9905-0125788.
PR 25-MAR-1999; 9905-0126264.
PR 29-MAR-1999; 9905-0126785.
PR 01-APR-1999; 9905-0127462.
PR 06-APR-1999; 9905-0128234.
PR 08-APR-1999; 9905-0128714.
PR 16-APR-1999; 9905-0129845.
PR 19-APR-1999; 9905-0130077.
PR 21-APR-1999; 9905-0130449.
PR 23-APR-1999; 9905-0130510.
PR 28-APR-1999; 9905-0130891.
PR 30-APR-1999; 9905-0132048.
PR 30-APR-1999; 9905-0132407.
PR 04-MAY-1999; 9905-0132484.
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 06-MAY-1999; 9905-0132487.
PR 07-MAY-1999; 9905-0132863.
PR 11-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.

PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144884.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148365.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149136.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.

PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.

PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 12.8%; Score 290; DB 21; Length 501;
Best Local Similarity 24.3%; Pred. No. 1,4e-15;
Matches 104; Conservative 84; Mismatches 196; Indels 44; Gaps 13;

QY 31 PKHESVTLKNGLOVSV-LENTGVLEVDVLYKVSRRNETMKSGIAHMLEHNRKSRK 89
DB 93 PETRYVTLPLRGVATVATESNLSAKTAT--GVWIDAGSRESDEFTNGTAHLEHMTFKTD 150
QY 90 NLKAGEFDKIVKRRGCVSNASTSFDTIRYIKTSQANLDSLELFAETGSLNLKEDEFL 149
DB 151 RRTVRALEEIEDIGHLANAYTSREQTYYAKYLDNVNQAOLDVLAIDLONSKFEQORIN 210
QY 150 PEROVVAEERR--WRDNSPIGMLYFRFNTAYVHPYHWPIDIGMDIIONTKLDIK 206
DB 211 RERDVILREMOEVEGQTEVDVLDHLH---ATAFOYTPRGRTILGPAQNVKSTIREDLON 266
QY 207 FHSLYYQPKNAIYLVGVDSOKVFEFLSKKHFESLKNLDEKAIPTPYMKERPODGARTAV 266
DB 267 YIKTHYTASRMVIAAGAVKHEEVVEQVKLFETKLS-DEPTTSQLVANEPASFTGSEVR 325
QY 267 VHKDGVHLEWVALGVKVPARKHKOVALDALSLRLEGE-----GKSSMLOSELVDK--RR 318
DB 326 MIDDDELPLAQFAVAFEGASWTDPDSVALVMQTMGLMSNMKNNGGKHVSDDLQGRAINE 385
QY 319 LNSOAFSHNMLODESVPLEFIAGGNPNVKAALQKEIVALLEKLYK--GEITQAEIDIKK 376
DB 386 IAESIMAFNTNYKDTGLF-----GVYAVAKADCDDLSVAIMVEVTKLAVRSPADVYTRAR 441
QY 377 INKADPISNLESSDVAGLIFADYLVONDI--QGLITPYQOF-----LD-LKVSIDLVR 426
DB 442 NOLKSSLLHMDGTSPIA-----EDIGRLTYGRIPTAELFARIDAVDAVSTVR 492
QY 427 VANEYFXD 434
DB 493 VANKIYD 500
RESULT 8
AAG35643

PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161922.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

Query Match 12.8%; Score 290; DB 21; Length 452;

Best Local Similarity 24.3%; Pred. No. 1,2e-15;
Matches 104; Conservative 84; Mismatches 196; Indels 44; Gaps 13;

QY 31 PKHESVTLKNGLOVYSP-LENTGVLEVDLYKVGSRNETMGSGIAHMLEHLEKSTK 89
DB 44 PETRVTLPLPGLRVATESNLAKTAT--VGWIDAGSRFESDEFNTGTAHFEHMFKGTD 101
QY 90 NLAGEBDKLVKRGGSVSNASTSPDIRRYFIKTSQANLDSLELFAETGSLNKEDEFL 149
DB 102 RRTYVRALEEEIEDIGHLNATYSREOTTYYAKYLDNSVNOALDYLAILONSKEFEQRIN 161
QY 150 PEROVVAEERR--WRTDNPISGLMYRFRNTAVYHYHWTPRIGFDDIQNTLKDIKK 206
DB 162 RERVDILREMGVEGQDEVDLHLH---ATAFYRPLGRTILGPAONVKSTIREDLON 217
QY 207 FHSIYYPRKNAIYLVGDVNSOKYFELSKKHESLKNLDEKAITPPMKPEKODGANTAV 266
DB 218 YIKHYIYASRNVIAAAGAVNHEEVQVKKLFKLSS-DEPTTSQLVANEPASFTGSEVR 276
QY 267 VHKDGVLEWVAGLYKYPARKHKOVALDLASRLGE---GKSSWLOSLEVDK--KR 318
DB 277 MIDDLDLPLAOPAVAFEGASMTDPDSVALMWQYMLGSMKNVGGKAVGSDLTORVAIN 336
QY 319 LASQAFSHNMOLODESEVLEFIAGSNPNVKAALOKETVALLLEKLR--GEITQAEIDKLK 376
DB 337 IASISIAFMNNYKDTGLF---GVYAAKADCDLDDLSYALIMEYETKLAYNSDADYRRAR 392
QY 377 INOKADTISNLESSDYAGLFADYLVONDI-QGLTDYOROF-----LD-LKVSPLVR 426
DB 393 NOLKSSLLHMDGTSPFA-----EDIGROLITFYGRIRIPTELPARIDAVASTVKKR 443
QY 427 VANEXFKD 434
DB 444 VANKTYTD 451

RESULT 7
AAG35644
AAG35644 standard; Protein; 501 AA.

AAG35644;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 43574.

KW Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

XX EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 9905-0121825.

PR 05-MAR-1999; 9905-0123180.
PR 09-MAR-1999; 9905-0123548.
PR 23-MAR-1999; 9905-0125788.
PR 25-MAR-1999; 9905-0126264.
PR 29-MAR-1999; 9905-0126785.
PR 01-APR-1999; 9905-0127462.
PR 06-APR-1999; 9905-0128234.
PR 08-APR-1999; 9905-0128714.
PR 16-APR-1999; 9905-0129845.
PR 19-APR-1999; 9905-0130077.
PR 21-APR-1999; 9905-0130449.
PR 23-APR-1999; 9905-0130510.
PR 23-APR-1999; 9905-0130891.
PR 28-APR-1999; 9905-0131449.
PR 30-APR-1999; 9905-0132048.
PR 30-APR-1999; 9905-0132407.
PR 04-MAY-1999; 9905-0132484.
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 06-MAY-1999; 9905-0132487.
PR 07-MAY-1999; 9905-0132863.
PR 11-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 14-MAY-1999; 9905-0134370.
PR 14-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 03-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 02-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.

PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-01343270.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142803.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.


```

XX Antigenic protein; vaccine; Lyme disease; infection; detection.
XX Borrelia burgdorferi.
XX OS
XX W09859071-A1.
XX
XX 30-DEC-1998.
XX
XX 18-JUN-1998; 98WO-US12718.
XX
XX 03-SEP-1997; 97US-0057483.
XX 20-JUN-1997; 97US-0050359.
XX 22-JUL-1997; 97US-0053344.
XX 22-JUL-1997; 97US-0053377.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (MEDI-) MEDIMUNE INC.
XX
XX Chai GH, Erwin AL, Hanson MS, Lathigra R;
XX
XX WPI: 1999-189980/16.
XX N-PSDB: AAX61711.
XX
XX New isolated Borrelia burgdorferi nucleic acids - used to develop
XX products for the diagnosis, prevention and treatment of diseases
XX caused by Borrelia, particularly Lyme disease
XX
XX Claim 12; Page 161; 275pp; English.
XX
XX This sequence represents a Borrelia burgdorferi (Bb) protein of the
XX invention, which is suitable for use in a vaccine. The Bb polypeptides
XX can be used in vaccines for eliciting protective antibodies to members of
XX the Borrelia genus, particularly for the use against Lyme disease in
XX humans and animals. They can be used for preventing or attenuating an
XX infection caused by a member of the Borrelia genus. The products can also
XX be used for detection of members of the Borrelia genus.
XX
XX Sequence 719 AA:
XX
XX Query Match 14.3%; Score 324.5; DB 20; Length 719;
XX Best Local Similarity 27.0%; Pred. No. 3.1e-18;
XX Matches 126; Conservative 95; Mismatches 181; Indels 65; Gaps 22;
XX
XX 1 MKHESYKRLGLSGVLTAVTGASMAQSYLPKHESVT--LKKGLQVVSPLKNGVIE 57
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 1 MNTQRIKANKCKPTSVLEFLFSCV--SNEIKLDQSLVKGKLVGLKRYIKKQTPKNAYN 58
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 58 VDLVLYKVGSRNEMTKSGIAHMLEHLNFKSTKRLKAGEEDKIVKRG--GVS-NASTSF 113
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 59 MGIWVNGSLNEDNERGIAHLEHMAFNGTDPGNSIVDLVKRGMQFGADINATSP 118
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 114 DITRYIKTSQAN---LKSLELFAETMGSLNKLDEFLPERQVVAEERRMTDNDSPIG 169
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 119 DFTYRRLDSDGNKDEIDESINILRNMAQSISFMKEEIDLENNIIEEK--KLGETYFG 176
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 170 MLY---FFRNTAYVYHPYHMTPIGFMDDIONMTLKIDKPHSLIYQPNAYLVVGDVN 226
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 177 RIYEKDKLTSGSLYE--FRSPIGLEQILSFQPEDFKFKYKMPRELASTIYVGGDID 234
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 227 SOKVELSKHESLKNLDEKALPPPYMKPEPKODGARTAVVHKD--VHLEWVALGYKVA 285
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 235 PIETEEKIKQFVSMKN-----PTDKIKEV---VSLDVELKDKLLEDELEVGPSLM 285
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 286 F-----KHKDOVALDALSR---LIGEGKSSWLOSELVDK--KRLASQAF---SHN 327
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 286 FKKKEIINFKTKRDL-LNAIKKSLLAALFENRFSSELKTAGVQFKNVSNKDFEFSKDN 344
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 328 MQLQDSVFLTAGGNPNKAKALQKEIYALLEKLKKGITQALDKLKNQKADISLU 387
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 345 NTIVAKSISL---NENPDHLNEGID--DFPELEIRIKKFGFTQGELEKV---RSQPKSL 396
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 388 E-----SSDVAGLFADLYQNDIGLFDY-QRQFLDKVSDLVAV 427
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

DB 397 ELRKNINKNTNSWALFQD-LIEIAINGSNKFPMNEXCDLSFOYLEKI 442
XX
XX RESULT 4
XX ID AAY20015 standard; Protein; 697 AA.
XX
XX AAY20015;
XX
XX 19-JUL-1999 (first entry)
XX
XX B. burgdorferi antigenic protein, t373.aa.
XX
XX Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
XX Borrelia burgdorferi.
XX
XX W09859071-A1.
XX
XX 30-DEC-1998.
XX
XX 18-JUN-1998; 98WO-US12718.
XX
XX 03-SEP-1997; 97US-0057483.
XX 20-JUN-1997; 97US-0050359.
XX 22-JUL-1997; 97US-0053344.
XX 22-JUL-1997; 97US-0053377.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (MEDI-) MEDIMUNE INC.
XX
XX Chai GH, Erwin AL, Hanson MS, Lathigra R;
XX
XX WPI: 1999-189980/16.
XX N-PSDB: AAX61712.
XX
XX New isolated Borrelia burgdorferi nucleic acids - used to develop
XX products for the diagnosis, prevention and treatment of diseases
XX caused by Borrelia, particularly Lyme disease
XX
XX Claim 12; Page 161; 275pp; English.
XX
XX This sequence represents a Borrelia burgdorferi (Bb) protein of the
XX invention, which is suitable for use in a vaccine. The Bb polypeptides
XX can be used in vaccines for eliciting protective antibodies to members of
XX the Borrelia genus, particularly for the use against Lyme disease in
XX humans and animals. They can be used for preventing or attenuating an
XX infection caused by a member of the Borrelia genus. The products can also
XX be used for detection of members of the Borrelia genus.
XX
XX Sequence 697 AA:
XX
XX Query Match 14.2%; Score 322; DB 20; Length 697;
XX Best Local Similarity 27.6%; Pred. No. 4.9e-18;
XX Matches 118; Conservative 87; Mismatches 162; Indels 60; Gaps 20;
XX
XX 38 LKNGLOVVSPLKNGVLEVDVLYKVGSRNEMTKSGIAHMLEHLNFKSTKRLKAGEED 97
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 17 LVNGLRYYIYKQTPKRNANMGIVFNWGLNEDNERGIAHLEHMAFNGTDPGNSIV 76
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 98 KIYKRG--GVS-NASTSFDIRYIKTSQAN---LDKSLFAETMGSLNKLDEFL 149
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 77 DVLKRFQMGFGADINATISFDFTYRRLDSDGNKDEIDESINILRNMAQSISFMKEE 136
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 150 PERQVVAEERRMTDNDSPIGMLY---FFRNTAYVYHPYHMTPIGFMDDIONMTLKIDK 206
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 137 LERNIIEEK--KLGETYGRIVYEKMDKFLTSGSLYE--FRSPIGLEQILSFQPEDFK 192
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 207 FHSLYQPNKNAIVLVGDVNSOKVELSKHESLKNLDEKALPPPYMKPEPKODGARTAV 266
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 193 FYRWKRPRLASTIYVGGIDPIETEEKIKQFVSMKN-----PTDKIKEV---VSLD 243
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
PS Claim 8; Page 395-398; 2054pp; English.
XX This sequence represents a Helicobacter pylori GMPD protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis.
XX
SQ Sequence 444 AA;
Query Match 100.0%; Score 2270; DB 19; Length 444;
Best Local Similarity 100.0%; Pred. No. 1,7e-180;
Matches 444; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKHFSVKRLGLISVLTGASMAHQSYLPKHESVTLKNGIQVYVPLENKTGVEVDY 60
Db 1 MKHFSVKRLGLISVLTGASMAHQSYLPKHESVTLKNGIQVYVPLENKTGVEVDY 60
QY 61 LYVGSRNEMTSGSIAMHLEHLNFKSTKNLKAGEFDKIVKRGVSNASTSFDTIRYFI 120
Db 61 LYVGSRNEMTSGSIAMHLEHLNFKSTKNLKAGEFDKIVKRGVSNASTSFDTIRYFI 120
QY 121 KTSQANLDKSLTEFETMGSINLKEDFELPERQVVAEERRRMTDPSPIGMLYFRFNAY 180
Db 121 KTSQANLDKSLTEFETMGSINLKEDFELPERQVVAEERRRMTDPSPIGMLYFRFNAY 180
QY 181 VYHPRHMTPIGFMDDIQMWTLLDKIKFHSIYQPKNAIVLVGDVNSQKVELSKHES 240
Db 181 VYHPRHMTPIGFMDDIQMWTLLDKIKFHSIYQPKNAIVLVGDVNSQKVELSKHES 240
QY 241 LKNLDEKAIPTPYKKEPKQDGAFTAVHKDGVHLEVALGYKVPAFKKHDOVALDALSLRL 300
Db 241 LKNLDEKAIPTPYKKEPKQDGAFTAVHKDGVHLEVALGYKVPAFKKHDOVALDALSLRL 300
QY 301 LGEKSSMLOSLELVKRRLASQAFSHNMQLODESVEFLIAGSNPVKAEALOKETIVALLE 360
Db 301 LGEKSSMLOSLELVKRRLASQAFSHNMQLODESVEFLIAGSNPVKAEALOKETIVALLE 360
QY 361 KKKKEITQAEIDKLIKQKADFIISNLESSSDVAGLFADYLVQNDIQGLTDYQROFLDK 420
Db 361 KKKKEITQAEIDKLIKQKADFIISNLESSSDVAGLFADYLVQNDIQGLTDYQROFLDK 420
QY 421 VSDLVVANEFKDTOSTTVFLKP 444
Db 421 VSDLVVANEFKDTOSTTVFLKP 444
RESULT 2
ID AAB62813 standard; Protein; 438 AA.
AC AAB62813;
XX
XX 02-MAY-2001 (first entry)
DE E. canis ProA protease amino acid sequence.
XX
XX Ehrlichia canis; dog; vaccine; canine monocytic ehrlichiosis; ProA;
KW protease A.
XX
XX Ehrlichia canis.
OS
XX
XX WO200107625-A2.
XX
XX 01-FEB-2001.
XX
XX 20-JUL-2000; 2000WO-US19763.
XX
XX

XX
PR 21-JUL-1999; 99US-0358322.
XX
XX (CORR) CORNELL RES FOUND INC.
PA
XX
XX Chang Y;
PI
XX
XX WPI; 2001-182797/18.
DR N-PSDB; AAF62017, AAF62021.
XX
PT Novel recombinant DNA that encodes a protein that elicits immune
PT response against Ehrlichia canis; useful for producing DNA vaccines for
PT protecting dogs against canine monocytic ehrlichiosis -
XX
XX Claim 3; Page 36-37; 45pp; English.
XX
XX This invention relates to DNA encoding a protein which elicits an immune
CC response against Ehrlichia canis. The DNA and protein encoded by it are
CC useful as components of a vaccine for treating E. canis infections in
CC dogs. The protein is useful for identifying a T-cell epitope against
CC E. canis. The vaccines are useful for protecting dogs against canine
CC monocytic ehrlichiosis. The present sequence represents ProA, a protease
CC from E. canis, which can be used to produce a vaccine to protect dogs
CC against canine monocytic ehrlichiosis.
XX
SQ Sequence 438 AA;
Query Match 16.3%; Score 370; DB 22; Length 438;
Best Local Similarity 25.7%; Pred. No. 2.5e-22;
Matches 112; Conservative 96; Mismatches 210; Indels 18; Gaps 10;
QY 17 LVTLGASMAHQSYLPKHESVTLKNGIQVYVPLENKTGVEVDYLYVGSRNEMTSGSGI 76
Db 10 IVFLTIANHALSFNKVYHEKLDNGMEYVYIPNHRAPAVMHA-VLYKVGSTDDPVGYSGL 68
QY 77 AHMLEHLNFKSTKNLKAGEFDKIVKRGVSNASTSFDTIRYFIKTSQANLDKSLTEFAE 136
Db 69 AHFEHLMFSGTE--KFPNLISTSNIGGNFNASTSQCTIYELIPROYLSLAMDIESD 126
QY 137 TMSGLNLEKEDFELPERQVVAEERRRMTDPSPIGMLYFRFNAYVYHMTPIGFMDDI 196
Db 127 RKQNKVYTDKALIRQKQVYVEERKKRVRVSOAKNILEEMEN-APYNGYGRPVYGMHEHI 185
QY 197 QNMTLKDIKFFHSIYQPKNAIVLVGDVNSQKVELSKHESLKNLDEKAIPTPYK- 255
Db 186 SNYNNEVAEAPFKLHYSNNAILIYTGADADPEVITLAKQYGIKPSNNK--PSSQYRV 243
QY 256 EPKQDGAFTAVHKDGVHLEVALGYKVP-AFKHDOVALDALSLRLEKSSMLOSLEYL 314
Db 244 EPPHKTNNMTLTKDSSVEIPELFLMYQIPNGITNKNYTLNMMLEILGSGKFSLYNDLV 303
QY 315 DKRRLASQAFSHNMQLODESVEFLIAGSNPN-VKAEALOKETIVALLEKKGETIQAEID 373
Db 304 INNPVITISIKTDYNYLTDSDNMTLSIEALPKNGISTEAEQETIHKICINNYLENGISAETLE 363
QY 374 KIKINQKADFIISNLESSSDVAGLFADYLV---QNDIQGLTDYQROFLD-LKVSQYVVA 428
Db 364 SAKYKVKAHILYAFDGLFISYFGMHLLGVPLEISNIYD---TIDKYSIDVNSAM 419
QY 429 NEYFKDTOSTTVFLKP 444
Db 420 ENIFONNIRLTGHLIP 435
RESULT 3
ID AAY20014 standard; Protein; 719 AA.
AC AAY20014;
XX
XX 19-JUL-1999 (first entry)
XX
XX B. burgdorferi antigenic protein, f373.aa.
XX
XX

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 11:53:19 ; Search time 40 seconds

(without alignments)
1479.082 Million cell updates/sec

Title: US-09-881-752a-212

Perfect score: 2270
Sequence: 1 MKHFSVKRLGLSLVLYTL.....VRVANEYFKDSTVFLKP 444

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : A_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the total score distribution, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2270	100.0	444	AAW98306	H. pylori GHPO 525
2	370	16.3	438	AAW98306	E. canis Proa prot
3	324.5	14.3	719	AAW20014	B. burgdorferi ant
4	322	14.2	697	AAW49837	Propionibacterium
5	312.5	13.8	355	AAW49837	Arbidopsis thalia
6	290	12.8	452	AAW35645	Arbidopsis thalia
7	290	12.8	501	AAW35644	Arbidopsis thalia
8	283.5	12.5	503	AAW35643	Arbidopsis thalia
9	283.5	12.5	438	AAW81200	Mycobacterium tube
10	262.5	11.6	941	AAW34459	Porphorymonas gling

11	262.5	11.6	945	20	AAW34335
12	260	11.5	427	23	ABP26991
13	237	10.4	460	21	AAW28098
14	237	10.4	462	21	AAW28097
15	236.5	10.4	429	23	ABP26992
16	236	10.4	415	20	AAW06398
17	234	10.3	489	21	AAW44658
18	234	10.3	489	22	AAW92545
19	233.5	10.3	445	23	ABP40103
20	233	10.3	489	20	AAW88114
21	233	10.3	489	21	AAW57558
22	231	10.2	428	23	ABW49448
23	228	10.0	434	22	AAW35305
24	219.5	9.7	1280	21	AAW48316
25	215.5	9.5	388	22	AAW34391
26	212.5	9.4	554	22	ABG19409
27	210	9.3	424	23	ABP40105
28	205.5	9.1	518	22	AAW80151
29	205.5	9.1	525	22	AAW79167
30	205.5	9.1	531	22	AAW23528
31	204	9.0	410	23	ABW55325
32	199	8.8	556	22	ABW58441
33	196.5	8.7	970	21	AAW31677
34	195.5	8.6	278	22	AAW48317
35	195	8.6	956	23	ABW93843
36	194.5	8.6	944	21	AAW31678
37	194	8.5	503	21	AAW17665
38	194	8.5	503	21	AAW43212
39	191.5	8.4	427	22	AAW38047
40	191	8.4	1059	21	AAW48317
41	190.5	8.4	427	22	AAW37850
42	189.5	8.3	446	21	AAW17666
43	189.5	8.3	446	21	AAW43213
44	188	8.3	1259	21	AAW18784
45	185.5	8.2	1019	22	AAW62531

ALIGNMENTS

RESULT 1	
AAW98306	
ID	AAW98306 standard; Protein; 444 AA.
AC	AAW98306;
XX	
DT	31-MAR-1999 (first entry)
XX	
DE	H. pylori GHPO 525 protein.
XX	
KW	GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis; peptic ulcer disease.
XX	
OS	Helicobacter pylori.
XX	
PN	WO9843478-A1.
XX	
PD	08-OCT-1998.
XX	
PF	01-APR-1998; 98WO-US06371.
XX	
PR	29-JUL-1997; 97US-0902615.
PR	01-APR-1997; 97US-0833457.
PR	24-JUN-1997; 97US-0881227.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
PA	(INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI	Al-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;
XX	
DR	WPI: 1998-542293/46.
DR	N-PSDB; AAX14025.
XX	

Porphorymonas gling
Streptococcus poly
Human secreted pro
Streptococcus poly
Bacillus subtilis
Human mitochondria
Human protein sequ
Staphylococcus epi
Mitochondrial proc
Human mitochondria
Listeria monocytog
Enterococcus faeca
Arbidopsis thalia
Staphylococcus aur
Novel human diagno
Staphylococcus epi
Human protein SEQ
Human protein SEQ
Novel human enzyme
Lactococcus lactis
Drosophila melanog
Arbidopsis thalia
S. epidermidis ope
Herbicidally activ
Arbidopsis thalia
Arbidopsis thalia
Arbidopsis thalia
Arbidopsis thalia
Streptococcus pneu
Arbidopsis thalia
Streptococcus pneu
Arbidopsis thalia
Arbidopsis thalia
A pea chloroplast
Ral insulin degrad

THIS PAGE BLANK (USPTO)

```

QY 96 FDKIYKRRGVSNNSTSDITRFFIKTSQANLMDKSLFELAFEMGSLNKKDEBEFLPERQV 155
Db 66 FDKIYKRRGVSNNSTSDITRFFIKTSQANLMDKSLFELAFEMGSLNKKDEBEFLPERQV 155
QY 156 AEEERRMTRDNPSPICMLYRFPFNFAVYHPYHWTPTGCFMDIDIONMTLKDKIFKHSLSYOPK 215
Db 126 LEEIRACVDD-PDWGFPCLQSOSTIYODHPYGRSVYGTGTEELMOQSPLEAMRRHRNHYOPE 184
QY 216 NAIYLVGVGDVNSQKAVFELSRRKHFESLKNLDE---KAIPTPYMKEPKODGARTAVVHKD 271
Db 185 NMTYVIAAGIAQQAAMELVNRSFENFSKRPVCCPLKVPKSPVTKGITHQELSLPRIEGAR 244
QY 272 VHLHEVALGITYVPAPKAKKDDVALDALSRLLBEGKSSWLOSELVYDKKRLASQAFSHNMLO 331
Db 245 LLMAWVAVGVE---QLRTAYGLDLISVYLAEGFRSTRLYVRDLREELQTV-OGICSNFSL 299
QY 332 DESAFELFIAGGNPNKAKALOKELVALLKELKKGITTOAELDKLINOKADPISLWESS 391
Db 300 CESSLEFVTAMLEPENLEQVEDLILSHLDDIQTSGVSRQOEIARRRLLCNEAFSTERNP 359
QY 392 DVAGLEADYLVONDIQGLDYQROFLDKSVDGLVNAEYFKDQSTVPLKP 444
Db 360 QLTGLIKGYNTIAQAEIAVTPPHQIQSFDQLOQLAKOHLSLONVAYTILKP 412

```

RESULT 15

zinc proteinaseX00816 [imported] - *Xylella fastidiosa* (strain 9a5c)
C:Species: *Xylella fastidiosa*
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: C82759
R:Anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: C82759
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1..990 <SIM>
A:Cross-references: GB:AE003921; GB:AE003849; NID:9105710; PIDN:AFR83626.1; GSPDB:GN001
R:Experimental source: strain 9a5c
R:Simpon, A.J.G.; Reinhard, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; P.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carver, F.
de Neto, E.; Docena, C.; El-Dorri, H.; Fachinani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohnt
J.D.; Jungueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Krumme, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Meick, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.
Rodrigues, V.; Rosa, A.C.R. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A:Authors: Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Contents: annotation
A:Genetics:
A:Gene: XF0816

Query Match	16.4%	Score 372.5	DB 2	Length 990
Best Local Similarity	29.3%	Pred. No. 6.4e-16		
Matches 125; Conservative	69	Mismatches 184	Indels 49	Gaps 15

```

0Y  9  LIGASSVLL-----VTIGASMHAASTYRK-----HESYTIKGLOVAYSPLFNKCVI  56
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  42  ILGAGTATFTPOVAAITGTLTISKAIKPAQIPDIAIERFTLPNGLVI--VHEDHKAPVI  1000
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y  57  EVDVLLKYGSSNETMGRSGIAHMLDEHINFKSTKNLKAQEEFDLYKRFGVGSNASTSDIT  116
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  101 AVNVTYHIGSDADPAGKTGFALHEHLMFSGSEHHKASYPQPEFKIGATENGNTWYDRT  160
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y  117 RYFKTQANIDKSLLELFAETMGSL--NLKEDDFLPEROVAEERRRRTDINSPIGMLYFR  174
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db	161	NVFOVPTTALDMLMMSDSBKMHLALAIQKELDTQRGVKNKRRQR - ENVPYGRATON	21.9
Oy	175	FENFNAV-YVHEHYHMTPIGCEMDDIONMTLKIDIKKHSLYOPKNAIVLVGDVNSOKYFEL	23.3
Db	220	ILSNLFPANHYOXSITGISMDELAASLADVKSFEOAHYGAANNITVLADIDIVAAEHRDK	27.9
Oy	234	SKHFEESLKNLDEKAIPTRYKE - PKOD-GARTAVVHRDGVHLEWVALGYKVPFAFKHDQ	29.1
Db	280	AAKFGODIPAGRPVAHQHPWITPLPAQKRGVQYQVQSPRLYRWI - - - - -TPELSGV	33.4
Oy	292	VALDALSLNLEGGSSWLOSLVYKKRLASO - - - - -AFSHMOQDESVEFLFIGGN	34.3
Db	335	VOLDIATITLIGKSSSRLOYRVLVYKDKLAISISAGISPFALASOMQINAO - - - - -VYPGI	36.9
Oy	344	PNVKAELAKEIVALLLEKLK - - - - -GETTOLELKLINOKADPISNLESSSDVAG - LFA	39.8
Db	390	DPAVEA - - - - -AIAEELKFKLEGP - SDELORAQMNYRSDPLVRELVGKKAAILAA	44.2
Oy	399	DLYVOND 405	
Db	443	GQLYIND 449	

Search completed: November 18, 2002, 11:59:06
Job time : 24 secs

Job time : 24 secs

Query Match 19.0%; Score 432; DB 2; Length 952;
Best Local Similarity 28.5%; Pred. No. 9.4e-20;
Matches 129; Conservative 81; Mismatches 182; Indels 60; Gaps 11;

9 LGLSSVILV---TLGASHMAOSYLPK-----HESYTLKNGLOVSV 47
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Db 7 LRGFSLLLAGCGSSSDSPSLPFESSLKPGVTLVEEVAEPKGYMPIPKSKYRLDGLVILIS 66
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Oy 48 PLENKGVIEVDVLYKSGRNETMKGSGIAHMLEHLNFKSTKNLKAAGEPFIKRGVYS 107
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Db 67 P-DDSPRLVHVDVTVHVGSAKEIGKSGFAHFEHMFQSGKIVGQOQHRLTEAGCSL 125
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Oy 108 NASTSFDIRYFIKTSQANLDSLELFAETMGS--NLKEDFLPEROVAEERKRTDN 165
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Db 126 NGTTNDRNTNYEFTVPANOLEKMLLEADRMGFLDAVSGRKFEIQRDVYKNNRAQNDYD 185
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Oy 166 SPIGMLYFFRFNTAYY--HPYHPTPIGFMDIONMTLKDKKHSLYQPKNAVLYVGD 224
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Db 186 RPYGLMWEKMGKALYDEGHPYSWOTIGYVSDLDKAVDNLKAFELRWYGNNAVLTIGGD 245
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Oy 225 VNSOKVFEISKHFEEL-KNLDEKALPTPYMKPEKODGARTAVVHKGVHLEWVALGYKV 283
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Db 246 LDVKQTLAWYQKTFKSGIPKGPVVDAP----KQPARLSEDRFTTLEDROYQPMILIGWPI 301
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Oy 284 PAFKHKDOVALDALSLRLGEGKSSWLOSELY-DKKRIASQAFSHNMLODESVFLFIAG 342
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Db 302 QYWGADVOVALDALASGKNSLLIYQELVKTQKAVDAGAFQDCALACTFVYVMAAPS 361
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Oy 343 NRPVKALELOKEIVALLEKIKKEITQAEIDKIKINOKADFIENLESSSVAGLPADYLY 402
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Db 362 GAKGKIAPLYQETLYQLEKFKQOGVASARLEQIIGSEASAVPALES-----408
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Oy 403 QNDIOGLTDYQROFLDKYSDLVAVNEXEKD 434
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Db 409 ---VKS-----KVSQI--AANOTFFD 424
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
RESULT 10
T35873
zinc protease - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: T35873
R:Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, November 1997
A:Reference number: 221592
A:Accession: T35873
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-450 <OLI>
A:Cross-references: EMBL:AL009204; PIDN:CAAL5794.1; GSPDB:GN00070; SCOEDB:SC9B10.04
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC9B10.04
C:Superfamily: mitochondrial processing peptidase alpha chain

Query Match 18.2%; Score 412.5; DB 2; Length 450;
Best Local Similarity 29.5%; Pred. No. 5.8e-19;
Matches 129; Conservative 76; Mismatches 175; Indels 57; Gaps 11;

38 LKNGLOVSVPLENKGVIEVDVLYKSGRNETMKGSGIAHMLEHLNFKSTKNLKAAGEPD 97
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Db 23 LKNGLRIV--LSEDLTPRAVAUCLMDVDSRHEVKGRTGLAHLEHLMFGSAGVKNQNH 81
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Oy 98 KLYKRGVSNASTSFDIRYFIKTSQANLDSLELFAETMGS--NLKEDFLPEROVY 155
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Db 82 ELVQGGGSLNGTSTSEERNTYETMPAHQLELALMLEADRMGSLAALDESMENQRDVY 141
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Oy 156 AEERKRTNSPIGLMYFFRFNTAYY--HPYHPTPIGFMDIONMTLKDKKHSLYY 212
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Db 142 KNERKRYNVPYGTAFKEL--TALAYPCHPHHTPTISMDALDAATLADAAAFRTYY 199
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Oy 213 QPKNAVLYVGDVNSOKVFEISKHFEELKNLDEKALPTPYMKPEKODGARTAV-----266
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
RESULT 11

processing proteinase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AE2446
R:kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuntz, T.; Sasamoto, S.; Watanabe, A.; Irigu
Nakasaki, N.; Shimpo, S.; Sugimoto, M.; Takasawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A>Title: Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE2446
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-426 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAW76824.1; PID:g17134263; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr5125
C:Superfamily: mitochondrial processing peptidase alpha chain

Query Match 17.8%; Score 403; DB 2; Length 426;
Best Local Similarity 27.0%; Pred. No. 2.2e-18;
Matches 113; Conservative 96; Mismatches 189; Indels 20; Gaps 9;

37 TLKNGLOV--SVPLENKGVIEVDVLYKSGRNETMKGSGIAHMLEHLNFKSTKNLKA 94
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Db 18 TLKNGLTIVEQMPVE---AVNLSLMDVGSSESDAINGMAHLEHMFKGTERTLAG 73
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Oy 95 EPDKTYKRGVSNASTSFDIRYFIKTSQANLDSLELFAETMGSNLKEDFLPEROV 154
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Db 74 EFERHIEERGAVTNAATISODTYHTYINTAPQDAKALPQIDVYVNASIPDEFERERV 133
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Oy 155 VAERKRTNSPIGLMYFFRFNTAYYVPHPTPIGFMDIONMTLKDKKHSLYYOR 214
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Db 134 VLEERK-RSEDNRRRTFRFAMETAQQLPYRPVLYGPRESSISQQLRQGMROFHASWYOR 192
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Oy 215 KNAVLYVGDVNSOKVFEISKHFEELKNLDEKALPTPYM--KEPKODGARTAVNKK---269
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Db 193 QSTTAAVAVNLEPEQDLEITTEGFFNQLKTPHSPPLTTPHSLHLEP---AFTEIYAREEV 248
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Oy 270 -DSVHLEWVALGYKQVAFRHNKDOV--ALDALSRLGSKSSWLOSELYVDKRLASQAFSHN 327
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Db 249 DESLQARLIMVWRVGLNQLQLEOTGLDVLGILAHGRTSRLYQDREERGLVTSISVN 308
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Oy 328 MLODESVEFLTAGNPNVKALELOKEIVALLEKIKKEITQAEIDKIKINOKADFI 387
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Db 309 MSNRLOGTF-YISAKAVEDLAAVEALIAQIHRIKLOTLEYLTELKELARVAKRANRPTFGN 367
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Oy 388 ESSSDVAGLPADY-LVQNDIOGLTDYQROFLDKYSDLVAVNEXEKDQSTTFVLKP 444
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
Db 368 ETPSDRAGLYGFGYQSLVGDLEPAFNYPRAHQTEAPDILLANQYLCPEAYGVVWKP 425
| | | | | : : : : : | | | | : : : : : | | | | : : : : :
RESULT 12

QY 443 KP 444
Db 455 LP 456

RESULT 7

S76140
hypothetical protein - *Synechocystis* sp. (strain PCC 6803)
C:Species: *Synechocystis* sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:Accession: S76140
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpe, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136 (1996)
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S76140
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-524 <K&N>
A:Cross-references: EMBL: D80914; GB: AB001339; NID: g1653477; PIDN: BAA18399.1; PID: d101913

* A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match

21.1%; Score 479; DB 2; Length 524;

Best Local Similarity 26.9%; Pred. No. 4e-23;
Matches 141; Conservative 88; Mismatches 186; Indels 110; Gaps 13;

QY 5 SVKRLDSSVLTGASM-----HAQSLP-----KHESVTLKNG 41
Db 8 SLHCLLVLTALILVHPGIMSHPAIADLAPTNNRNSQTLTPYLNRALRIORITEFOLDNG 67
QY 42 LQVAVSVLENKGVIEVDVLYKVSNNETMGSGIAHLEHNFKSTKN----- 90
Db 68 LKFI-VHNNENAPVSVFYTYEDVGVDPEPKGTGAHFLHLEMARKGTERIGTKPTOEQQ 126
QY 91 -----LKAEEFKIVKREGVS-N 108
Db 127 LLDLDQVFAQITPARAKGDKTGEQKIQEQFKQIQOQAQDLIKONEFGQIOMAGVGLN 186
QY 109 ASTSFDTIRFIKTSQANLQKSLLEFAETMSLMLKDEDFEPQOVAAEERKMTDINSPI 168
Db 187 AATSDATFFYFSLPSKLELMMSLESEF--LEPVREFYQEQEVLLEERMTENNVP 244
QY 169 GMLYFREFNTAYVYHPYHMTPIGEMDDIQMTLKDKFHSLYQPKNAIYLVYGDVNSQ 228
Db 245 GQWVEELDFRAFTKHPRRPVITGDEDIRLSRODVIDFEKTIIPGNMTIAYVGVKVD 304
QY 229 KVFELSKKHESLKNLDEKAIPTPYMK--EPKODGARTAVVHKDGVH-----EMVALGY 281
Db 305 QVKSIAQKFGFR-----PQRPPTQPVTVVEBPQ-----TQKEINLTLPSQWYFEGY 353
QY 282 KVPFKHKDOVALDALSRLGEGSSWLOSSELYDKRLASQAFSHNMOLDE--SVLFIT 339
Db 354 HSPFEDDPDSAVFVMTTILSSGTSRLYOSLVEEKLALMAQGFNFPADKFPNLLMFY 413
QY 340 AGGNPNVKAELQKEIYALLEKLEKGETQAEIDKLKINOKADISNLESSDVAGFEAD 399
Db 414 AQSPAGSLDLSALHGEIERLKMPEVTPPELEERAQNLQTSALQSLNSMGMAQLLYK 473
QY 400 YLVONDIQGLTYQROFLDKV-----SDLVVANEYFKDTOST 438
Db 474 YNVRTG-----DMRNLFARLEAIAVTPEDIQVFAQETFEFRENST 513

RESULT 8

D83598
Probable zinc proteinase PA0372 [imported] - *Pseudomonas aeruginosa* (strain PA01)
C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83598
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000

A:Title: Complete genome sequence of *Pseudomonas aeruginosa*, PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: D83598
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-465 <STO>

A:Cross-references: GB: AE004475; GB: AE004091; NID: g9946221; PIDN: AAG03761.1; GSPDB: GN
A:Experimental source: strain PA01

C:Genetics:
A:Gene: PA0372

Query Match

19.7%; Score 447; DB 2; Length 465;

Best Local Similarity 29.4%; Pred. No. 3.8e-21;
Matches 136; Conservative 96; Mismatches 183; Indels 48; Gaps 15;

QY 7 KRLIGSSVLTGASMAQS--YLPKHESVTLKNGIQVAVSVLENKGVIEVDVLYKY 64
Db 15 RRRVGL---LLASCLPLFQAQAEQTPHE-FSLDNGLKVI-VREDHRAVYVSQLMYRI 69
QY 65 GSRNETGKSGIAHLEHNFKSTKNKAGFEKIVKRGVSNASTSFDTIRFIKTSQ 124
Db 70 GSSYETGELTGLSHALHMMFKGSRKLGPGASRVLDLAEBNAFTTDDYATAYVGLAR 129
QY 125 ANLDKSELFAETGSLNKLKEDFLPEROVAAEERKMTDINSPIGMLYFPFNTAYVYHP 184
Db 130 DRLVVALEMEADRNAHSLPVDQKSELEYEKERRIRTDNPAALFEFKAAAYASG 189
QY 185 YHMTPIGEMDDIQMTLKDKFHSLYQPKNAIYLVYGDVNSQVLSKKEHSL--K 242
Db 190 YHPTPIGMADLQMTIDDLHMYESWYAPNNATLVYGDVTADEVTKLAKRYGEIPIWR 249
QY 243 NLDEKAIPTPYMKRPPKODGARTAVVHNDGVHLEVALGYKVPAP---KKHDOVALDALS 298
Db 250 QLPAPRRPLE-LAEPGRRLKLVY---RTQLPNLIMGFVNPISGSSSENPREVALRLIG 304
QY 299 RLLEGKSSWLOSSELYDKRLASQAFSHNMOLDESVLFLAGNPNV-KAELQKEIVA 357
Db 305 ALLDGGYSARLASLBERGEELVAGASTY-YDAFNRGSLVLSKTPVQKGTLEQVAG 363
QY 358 L--LEKIKRGETQAEIDKLK-----INQKADISNLESSDVAGLEADYL 401
Db 364 LMKQDLQKQPPSAELIERVRAQIMAGVYKESDIAQSSISQLES---VGL-SMKT 418
QY 402 VQNDIQGLTYQROFLDKVSDLVVANEYFKDTOSTVLEKP 444
Db 419 IDQLEALK-----AVTPDDIQKARTYFTPSRLTLAQLVP 454

RESULT 9

C82309
Probable insulinase-type proteinase VC0554 [imported] - *Vibrio cholerae* (strain N1696

C:Species: *Vibrio cholerae*
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C:Accession: C82309
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000

A:Title: DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: C82309
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-952 <HEI>

A:Cross-references: GB: AE004141; GB: AE003852; NID: g9654976; PIDN: AAF93722.1; GSPDB: GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:
A:Gene: VC0554
A:Map position: 1

```

Db      2 AQAALPELSRLDGVSNFLTPNGMOVVYIP-DHRAVYVOTMWYHVGADDEARPGVSGIAHF 60
QY      80 LEHNFKSTKNLKAGEFDKIVKREGVSNASTSPDITRFFYTKTSQANDKSLTEPAETMG 139
Db      61 LEHLMFKGTKNHHPAGEFSARIASIGGQNAFTSYDYTAQVQVSPSEALEMWDPESDRME 120
QY      140 SLNKEDEFLPEBROVVAEERKRRDINSIGMIYFFEEFATVYHNYHHTPRIGEMDIONW 199
Db      121 NLVDEEVRVATEREYLTDERRMRIDNSNGAMIMENTDAVLEYNHPRRPVIGMOQEBKL 180
QY      200 TLKDKKKRHSLYXQPKNALVWGVGNQOKXPELSKKAFESJLNDEKAIPTPYKKBEKQ 259
Db      181 SLKNADIFYNOYYPNNATLVLAGDYTERVERELAMKMANVHKRAE-VLLRERPOEPAK 239
QY      260 DGARTAAVHKDGVHLEWALGYKVAFA-----KHKDQVADALPSRLLGEGSKWLD 310
Db      240 HAARVYTLHDERVSTPSPRISMVLPSTANYEKRFANVKKGDAPALDLLEITIGSOLSKLY 299
QY      311 SELVDKRLKASQ-AESHMLODESVPFLFIAGNPBNVKAALOKEIVALLLEKTKKEITQ 369
Db      300 QQLIYKQCIAMETGASVGDALDDETFESVYVGPBNGASISLGEKAVANAQVRIINDGTQ 359
QY      370 AELDKLKTNQKADFTSNLESSDVAGLEADYL---VONDIOGLTDYORQFLDKVSDLV 425
Db      360 AELDQARRRFLKAVIFARDSOTGMARIGSALSVSQTVDDIOKWPDLIK---SVTVQDIK 416
QY      426 RVANFEYKDJOSTYVEFKP 444
Db      417 DVARRYLVKDAQVTSYLLP 435

```

RESULT 5
AF2048
proteinase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AF2048
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuriitz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759640
A:Accession: AF2048
A:Stratus: preliminary
A:Molecule type: DNA
A:Residues: 1-528 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA073639.1; PID:917131030; GSEDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
C:Gene: all1940

[illegible][illegible]

```

RESULT 6
F87693
peptidase, M16 family [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequenceRevision 20-Apr-2001 #textChange 20-Apr-2001
C:Accession: F87693
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; J. Labb, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: F87693
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-948 <SMO>
A:Cross-references: GB:AE005673; NID:g13425328; PIDN:AAK25546.1; GSPDB:GN00148
C:Genetics:
C:Gene: CC3584

```

[illegible]

QY 421 VSDLVRYANEFKDTQSTTFVFLKP 444
 |||||||||||||||||||||
 Db 421 VSDLVRYANEFKDTQSTTFVFLKP 444

 RESULT 2
 D71935
 Probable zinc proteinase - Helicobacter pylori (strain J99)
 C:Species: Helicobacter pylori
 A:Variety: strain J99
 C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 17-Mar-2000
 C:Accession: D71935
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
 Nature 397, 176-180, 1999
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric patho-
 A:Reference number: A71800; MUID:99120557; PMID:9923682
 A:Accession: D71935
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-443 <ARN>
 A:Cross-references: GB:AE001475; GB:AE001439; NID:g4154939; PIDN:AAD05993.1; PID:g4154939
 A:Experimental source: strain J99
 C:Genetics:
 A:Gene: jhp0411
 C:Superfamily: mitochondrial processing peptidase alpha chain

 Query Match 94.7%; Score 2149.5; DB 2; Length 443;
 Best Local Similarity 95.7%; Pred. No. 4, 2e-130;
 Matches 425; Conservative 7; Mismatches 11; Indels 1; Gaps 1;

 QY 1 MKHRSVKRLGLSSVLLVLTGLASMAQSYLPKHESVTLKNGLOVSVPLENKTGVEVDV 60
 |||||||||||||||||||||
 Db 1 MKHRSVKRLRLSSVLLVLTGLASMAQSYLPKHESVTLKNGLOVSVPLENKTGVEVDV 60

 QY 61 LYKYGSRNETWGKSGIAHMLEHLNFKSTPKNLKAGEFDKIVKRGGSVMASTSDITRYFI 120
 |||||||||||||||||||||
 Db 61 LYKYGSRNEVWGKSGIAHMLEHLNFKSTPKNLKAGEFDKIVKRGGSVMASTSDITRYFI 120

 QY 121 KTSQANLDKSLLEFAETGSLNLKEDFLPEROVAAEERRMRTDNPIGMLYFRFFNTAY 180
 |||||||||||||||||||||
 Db 121 KTSQANLDKSLLEFAETGSLNLKEDFLPEROVAAEERRMRTDNPIGMLYFRFFNTAY 180

 QY 181 VYHRYHNTPIGFMDIQWTLKDIKKFSLSLYQPKAIVLVGDVNSQKYFELSKHFEES 240
 |||||||||||||||||||||
 Db 181 STTP-TMDAIGFMIDIQWTLKDIKKFSLSLYQPKAIVLVGDVNSQKYFELSKHFEES 239

 QY 241 LKNLDEKAIPPYKKPEPKDQAGARTAVVHRDGVHLEVALGYKVPARKHKQVALDALSL 300
 |||||||||||||||||||||
 Db 240 LKNLDEKAIPPYKKPEPKDQAGARTAVVHRDGVHLEVALGYKVPARKHKQVALDALSKL 299

 QY 301 LGEKSSWLOSELYDKKRLASQAFSHNMQLODESVFLTAGGNPNYKAALQKEIVALLE 360
 |||||||||||||||||||||
 Db 300 LGEKSSWLOSELYDKKRLASQAFSHNMQLODESVFLTAGGNPNYKAALQKEIVALLE 359

 QY 361 KLRGGETQAEALDKIKIQKADFISNLESSDVAGFAFYIVONDIOGLLDYOROFDLK 420
 |||||||||||||||||||||
 Db 360 KLRGGETQAEALDKIKIQKADFISNLESSDVAGFAFYIVONDIOGLLDYOROFDLK 419

 QY 421 VSDLVRYANEFKDTQSTTFVFLKP 444
 |||||||||||||||||||||
 Db 420 VSDLVRYANEFKDTQSTTFVFLKP 443

 RESULT 3
 E81352
 Probable zinc proteinase Cj0805 [Imported] - Campylobacter jejuni (strain NCTC 11168)
 C:Species: Campylobacter jejuni
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
 C:Accession: E81352
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillingworth, T.;
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrett, T.;

Nature 403, 665-668, 2000
A.Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals A.Reference number: AB1250; MUID:20150912; PMID:1068204
A.Accession: EBI352
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-416 <PAR>
A.Cross-references: GB:AII3076; GB:AII1168; MID:g6968128; PIDN:CAB73070.1; PID:g696
A.Experimental source: serotype O2, strain NCTC 11618
C.Genetics:
A.Gene: Cj0805
C.Superfamily: mitochondrial processing peptidase alpha chain

Query Match 55.1%, Score 1251.5; DB 2; Length 416;
Best Local Similarity 57.2%; Pred. No. 1.le-72;
Matches 235; Conservative 75; Mismatches 98; Indels 3; Gaps 2;

OY 33 HESVTLKNGLVSVPLDNKTGVLEVDLYKVGSNNEMTSGIAHMLHELNFKSTNLK 92
 :::|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 4 YEKILKKLEYALAPVKNSDVISDIIFYRVSNNELMGSSGTAHHMLELNFKSTNLK 63

OY 93 AGEPEIKYRFPGCVSNASTSPDIRFYFRTSQANIDKSLEFAETMGLNKDEEFLPER 152
 |::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 64 AGEPEIYKRGEGVNASTGFDTHTTYTKCAKNLDKLAEPAELMANLINLKDEFQPER 123

OY 153 OVVAEEERMRWDNSPDMGYFRFFNTAYVHYHPWTPIGFMDDIONMTLKDKIFHSLLY 212
 L24 AVLEEERRWRDNDNFLGIYLFRLFNHAFMYPHYWTPIGFPKDIEENSIEDIKFHSITY 183
DB

OY 213 QPKNAIVLVGDVNSQAKEFEISKHFESLNIDEKAIPPMKEBKODGAATAVYHKDG V 272
 ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 184 QPKNALILLVGISIESEVEFELSCKHFEKIKN-TPTLPKHTEPKODGVARRYLHNKS - 240

OY 273 HEWMVALCYKPVPAPHKOVALDALSRLLGGSKSSLÖSELVDDKKRLASQAFSHMLOD 332
 :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 241 DETELLALKRPINFNHAKIIPALNALSELLSGKSLSEIIDUNLTINDYRAYVNCID 300

OY 333 ESVFPLIAGNPNVAAEAIOKEITVALLEKLRKGETTQAEIDKLIKNOADPFISMLESSSD 392
 |:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 301 ENLFFICNPCPNVAAEKEVEKELLIIDKIMKGRISOQDIDRVANNVXSDTFISLNASA 360

OY 393 VAGEPADVLVONDIOGTDLDYOROFIDLKVSULVRVANEFKYOSTTVYEK 443
 ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 361 VANIGSYLARODINPLLNEYEDIONELKDLISCARKXYFOENSTVIIR 411

RESULT 4
ADJ639 zinc proteinase (EC 3.4.99.-) [imported] - Brucella melitensis (strain 16m)
C.Species: Brucella melitensis
C.Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C.Accession: ADJ639
R.DeVelvecchio, V.G.; Kaparal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov
. Mazur, M.; Goldsmann, E.; Selkov, E.; Elzezer, P.H.; Hagius, S.; O'Callaghan, D.; Let
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A>Title: The genome sequence of the facultative intracellular pathogen Brucella melit
A.Reference number: ADJ252; PMID:11756688
A.Accession: ADJ639
A>Status: Preliminary
A.Molecule type: DNA
A.Residues: 1-464 <KUR>
A.Cross-references: GB:AE008918; PIDN:RAL54279.1; PID:g17985255; GSPDB:GN00191
A.Experimental source: strain 16m
C.Genetics:
A.Gene: BME11037
A.Map position: II
C.Keywords: hydrolase

Query Match 22.4% ; Score 507.5; DB 2; Length 464;
Best Local Similarity 30.8%; Pred. NO. 5.1e-25;
Matches 135; Conservative 85; Mismatches 194; Indels 25; Gaps . 7;

OY 26 AOSYLPR-----HESVTLNGLQVSVPLDNKTGVLEVDLYKVGSNNEMTGSGIAAHM 79

THIS PAGE BLANK (USPTO)

FILING DATE: 06-JUN-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Robert Berliner
 REGISTRATION NUMBER: 20,121
 REFERENCE/DOCKET NUMBER: 1920-305D2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 213/977-1001
 TELEFAX: 213/977-1003
 INFORMATION FOR SEQ. ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 783 base pairs
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-465-995A-6

Query Match 1.6%; Score 7; DB 1; Length 783;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 300 LIGEGKS 306
 |||||
 Db 300 LIGEGKS 306

Search completed: November 18, 2002, 12:05:19
 Job time : 21 secs

ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/756,317
FILING DATE: 25-NOV-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/007,693
FILING DATE: 29-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Patterson, Melinda L.
REGISTRATION NUMBER: 33,062
REFERENCE/DOCKET NUMBER: MOBT:008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (713) 787-1400
TELEFAX: (713) 787-1440
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 560 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-756-317-11

Query Match 1.6%; Score 7; DB 2; Length 560;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 LKNGLOV 44
|||||
Db 244 LKNGLOV 250

RESULT 13
US-09-385-742B-6
; Sequence 6, Application US/09385742B
; Patent No. 6391611
; GENERAL INFORMATION:
; APPLICANT: Doi, Yoshiharu
; APPLICANT: Fukui, Toshiaki
; APPLICANT: Matsusaki, Hiromi
; TITLE OF INVENTION: POLYESTER SYNTHASE AND A GENE CODING FOR
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 07898-04/001
; CURRENT APPLICATION NUMBER: US/09/385,742B
; CURRENT FILING DATE: 1999-08-30
; PRIOR APPLICATION NUMBER: JP 82965/1997
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Pseudomonas
US-09-385-742B-6

Query Match 1.6%; Score 7; DB 4; Length 560;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 LKNGLOV 44
|||||
Db 244 LKNGLOV 250

RESULT 14
US-08-101-593-6
; Sequence 6, Application US/08101593
; Patent No. 5547859
; GENERAL INFORMATION:
; APPLICANT: Goodman, Myron F.
; APPLICANT: Rena-Krantz, Linda J.
; TITLE OF INVENTION: NEW DNA SEQUENCING ENZYMES
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Robbins, Berliner & Carspm
; STREET: 201 No. 5547859th Figueroa Street, Fifth Floor
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90012-2628
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/101,593
FILING DATE: 19930802
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Spitals, John P.
REGISTRATION NUMBER: 29,215
REFERENCE/DOCKET NUMBER: 1920-305
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 977-1001
TELEFAX: (213) 977-1003
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 783 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-101-593-6

Query Match 1.6%; Score 7; DB 1; Length 783;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 300 LLGEGKS 306
|||||
Db 300 LLGEGKS 306

RESULT 15
US-08-465-995A-6
; Sequence 6, Application US/08465995A
; Patent No. 5660980
; GENERAL INFORMATION:
; APPLICANT: Myron F. Goodman
; APPLICANT: Linda J. Rena-Krantz
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AND ISOLATING
; TITLE OF INVENTION: VARIANT T4 POLYMERASES
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Robbins, Berliner & Carson, LLP
; STREET: 201 No. 5660980th Figueroa Street, Suite 500
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90012-2628
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: ASCII DOS/TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,995A

Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0;
OY 103 FGVSNNA 109
DB 373 FGVSNNA 379

RESULT 10
US-09-072-596-106
; Sequence 106, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillion, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: MAKI, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-072-596-106

Query Match 1.6%; Score 7; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0;
OY 103 FGVSNNA 109
DB 373 FGVSNNA 379

RESULT 11
US-09-171-337A-5
; Sequence 5, Application US/09171337A
; Patent No. 6300095
; GENERAL INFORMATION:
; APPLICANT: BARRERO FUENTE, Jose Luis
; APPLICANT: RODRIGUEZ SAIZ, Marta
; APPLICANT: COLLADOS DE LA VIEJA, Alfonso J.
; APPLICANT: MORENO VALLE, Miguel Angel

SALTO MALDONADO, Francisco
DIEZ GARCIA, Bruno
TITLE OF INVENTION: PROMOTERS OF THE GENES GLUTAMATE
DESHYDROGENASE, -N-ACETYLHEXOSAMINIDASE
AND -ACTIN AND THEIR USE IN FILAMENTOUS
FUNGI EXPRESSION, SECRETION AND ANTISENSE

NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: LADAS & PARRY
STREET: 26 WEST 61 STREET
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10023
COMPUTER READABLE FORM:
MEDIUM TYPE: 3-1/4" Disk 1.44MB
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Microsoft Windows for Workgroups 3.11
SOFTWARE: Wordperfect 8 for Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/171,337A
FILING DATE: 14-May-1999
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/ES98/00056
FILING DATE: 5-MAR-1998
APPLICATION NUMBER: ES9700482
FILING DATE: 5-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: MASS, Clifford J.
REGISTRATION NUMBER: 30,086
TELECOMMUNICATION INFORMATION:
TELEPHONE: 233288
INFORMATION FOR SEQ ID NO: 5
SEQUENCE CHARACTERISTICS:
LENGTH: 461 amino acids
TYPE: amino acids
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Penicillium chrysogenum
FEATURE:
OTHER INFORMATION: amino acid sequence of the
glutamate dehydrogenase enzyme
(EC.1.4.1.4) with a molecular weight
of 49837 Da.
SEQUENCE DESCRIPTION: SEQ ID NO: 5
US-09-171-337A-5

Query Match 1.6%; Score 7; DB 4; Length 461;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0;
OY 42 LQVVSVP 48
DB 38 LQVVSVP 44

RESULT 12
US-08-756-317-11
; Sequence 11, Application US/08756317
; Patent No. 5849894
; GENERAL INFORMATION:
; APPLICANT: Clemente, Thomas E.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Mitsky, Timothy A.
; APPLICANT: Stark, David M.
; TITLE OF INVENTION: Improved Rhodospirillum Rubrum
TITLE OF INVENTION: Poly-B-Hydroxyalkanoate synthase
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 111:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-818-112-111

Query Match 1.6%; Score 7; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 FGVSN 109
Db 373 FGVSN 379

RESULT 8
US-08-818-111-106
Sequence 106, Application US/08818111
Patent No. 633852
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 106:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-818-111-106

Query Match 1.6%; Score 7; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 FGVSN 109
Db 373 FGVSN 379

RESULT 9
US-09-056-556-111
Sequence 111, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 111:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-056-556-111

Query Match 1.6%; Score 7; DB 4; Length 396;

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 130:
US-09-115-737-130

Query Match 1.6%; Score 7; DB 4; Length 11;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 358 LLEKLK 364
|||||
DB 4 LLEKLK 10

RESULT 5
US-08-233-146-2
Sequence 2, Application US/08233146
Patent No. 5648256
GENERAL INFORMATION:
APPLICANT: BEPPU, TERUHIKO
APPLICANT: YAMADA, HIDEAKI
APPLICANT: NAGASAWA, TORU
APPLICANT: HORINOUCHI, SUEHARU
APPLICANT: NISHIYAMA, MAKOTO
TITLE OF INVENTION: A GENE ENCODING A POLYPEPTIDE HAVING
TITLE OF INVENTION: NITRILE HYDRATASE ACTIVITY, A TRANSFORMANT CONTAINING THE
TITLE OF INVENTION: GENE AND A PROCESS FOR THE PRODUCTION OF AMIDES USING THE
NUMBER OF INVENTION: 8
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: N.Y.
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,146
FILING DATE: 22-APR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/694,746
FILING DATE: 02-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: MISROCK, S. LESLIE
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7005-024-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 220 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Pseudomonas chlororaphis
STRAIN: B23 (FERM BP-187)
US-08-233-146-2

Query Match 1.6%; Score 7; DB 1; Length 220;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 367 ITQAEID 373

DB 91 ITQAEID 97
|||||

RESULT 6
US-08-463-470-2
Sequence 2, Application US/08463470
Patent No. 5789211
GENERAL INFORMATION:
APPLICANT: BEPPU, TERUHIKO
APPLICANT: YAMADA, HIDEAKI
APPLICANT: NAGASAWA, TORU
APPLICANT: HORINOUCHI, SUEHARU
APPLICANT: NISHIYAMA, MAKOTO
TITLE OF INVENTION: A GENE ENCODING A POLYPEPTIDE HAVING
TITLE OF INVENTION: NITRILE HYDRATASE ACTIVITY, A TRANSFORMANT CONTAINING THE
TITLE OF INVENTION: GENE AND A PROCESS FOR THE PRODUCTION OF AMIDES USING THE
NUMBER OF INVENTION: 8
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: N.Y.
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,470
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/233,146
FILING DATE: 22-APR-1994
APPLICATION NUMBER: US 07/694,746
FILING DATE: 02-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: MISROCK, S. LESLIE
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7005-024-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 220 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Pseudomonas chlororaphis
STRAIN: B23 (FERM BP-187)
US-08-463-470-2

Query Match 1.6%; Score 7; DB 1; Length 220;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 367 ITQAEID 373
|||||
DB 91 ITQAEID 97

RESULT 7
US-08-818-112-111
Sequence 111, Application US/08818112
Patent No. 6290969
GENERAL INFORMATION:

STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: INDIANA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/641,318
FILING DATE: 18-AUG-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/816,178A
FILING DATE: 12-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: GAYLO, PAUL J.
REGISTRATION NUMBER: 36,808
REFERENCE/DOCKET NUMBER: X-10579
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0756
TELEX: (317) 276-3861
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 912 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-641-318-2
Query Match 1.8%; Score 8; DB 4; Length 912;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 92 KAGEFDKI 99
|||||
Db 261 KAGEFDKI 268
RESULT 3
US-08-465-325-130
Sequence 130, Application US/08465325
Patent No. 5686563
GENERAL INFORMATION:
APPLICANT: Magalini Pharmaceuticals Inc.
APPLICANT: 5110 Campus Drive
APPLICANT: Plymouth Meeting, PA 19462
TITLE OF INVENTION: Biologically Active Peptides Having
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I. Street, N.W. Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,325
FILING DATE: 05-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/184,462
FILING DATE: 18-JAN-94
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/891,201
FILING DATE: 01-JUN-92
ATTORNEY/AGENT INFORMATION:
NAME: Fordis, Jean B
REGISTRATION NUMBER: 32,984
REFERENCE/DOCKET NUMBER: 05387.0021-03000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 130:
SEQUENCE CHARACTERISTICS:
LENGTH: 11 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-465-325-130
Query Match 1.6%; Score 7; DB 1; Length 11;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 358 LLEKLK 364
|||||
Db 4 LLEKLK 10
RESULT 4
US-09-115-737-130
Sequence 130, Application US/09115737
Patent No. 6348445
GENERAL INFORMATION:
APPLICANT: U. Prasad Karl
APPLICANT: Taffy J. Williams
APPLICANT: Michael McNamee
TITLE OF INVENTION: Biologically Active Peptides With Reduced
TOXICITY IN ANIMALS AND A METHOD FOR PREPARING SAME
NUMBER OF SEQUENCES: 156
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner, L.L.P.
STREET: 1300 I Street, N.W. Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.3
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/115,737
FILING DATE: 15-JUL-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/465,330
FILING DATE: 05-JUN-1995
APPLICATION NUMBER: 08/184,462
FILING DATE: 18-JAN-94
APPLICATION NUMBER: 07/891,201
FILING DATE: 01-JUN-92
ATTORNEY/AGENT INFORMATION:
NAME: Fordis, Jean B
REGISTRATION NUMBER: 32,984
REFERENCE/DOCKET NUMBER: 05387.0021-06000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 130:
SEQUENCE CHARACTERISTICS:
LENGTH: 11 amino acids
TYPE: amino acid

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 12:02:40 ; Search time 18 Seconds
(without alignments)
725.765 Million cell updates/sec

Title: US-09-881-752a-212

Perfect score: 444

Sequence: 1 MKHFSVKRLGLSLVLT...VRANEYFKDQSTVFLKP 444

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/pdata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/pdata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/pdata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/pdata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/pdata/1/1aa/CTOS.COMB.pep:*
6: /cgn2_6/pdata/1/1aa/Dackfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	1.8	912	US-08-617-785-2	Sequence 2, Appl1
2	8	1.8	912	US-09-641-318-2	Sequence 2, Appl1
3	7	1.6	115	US-08-465-325-130	Sequence 130, App
4	7	1.6	114	US-09-115-737-130	Sequence 130, App
5	7	1.6	220	US-08-233-146-2	Sequence 2, Appl1
6	7	1.6	220	US-08-463-470-2	Sequence 2, Appl1
7	7	1.6	396	US-08-818-112-111	Sequence 111, App
8	7	1.6	396	US-08-818-111-106	Sequence 106, App
9	7	1.6	396	US-09-056-556-111	Sequence 111, App
10	7	1.6	396	US-09-072-596-106	Sequence 106, App
11	7	1.6	461	US-09-171-337A-5	Sequence 5, Appl1
12	7	1.6	560	US-08-756-317-11	Sequence 11, Appl1
13	7	1.6	560	US-09-385-742B-6	Sequence 6, Appl1
14	7	1.6	783	US-08-101-593-6	Sequence 6, Appl1
15	7	1.6	783	US-08-465-995A-6	Sequence 6, Appl1
16	7	1.6	783	US-08-465-994C-6	Sequence 6, Appl1
17	7	1.6	3248	US-08-353-700-1	Sequence 1, Appl1
18	7	1.6	3248	PCR-US95-16216-1	Sequence 1, Appl1
19	6	1.4	6	US-08-374-560-28	Sequence 28, Appl1
20	6	1.4	12	US-08-836-047-16	Sequence 16, App
21	6	1.4	16	US-08-940-095-254	Sequence 254, App
22	6	1.4	16	US-08-940-093-254	Sequence 254, App
23	6	1.4	16	US-08-940-096-254	Sequence 254, App
24	6	1.4	16	US-09-465-719-254	Sequence 254, App
25	6	1.4	16	US-09-453-605-254	Sequence 254, App
26	6	1.4	16	US-09-453-838-254	Sequence 254, App
27	6	1.4	21	US-09-391-799-4	Sequence 4, Appl1

28	6	1.4	22	1	US-07-928-930A-4	Sequence 4, Appl1
29	6	1.4	22	1	US-08-288-568-4	Sequence 4, Appl1
30	6	1.4	22	1	US-08-487-461-4	Sequence 4, Appl1
31	6	1.4	22	1	US-08-432-691-4	Sequence 4, Appl1
32	6	1.4	22	1	US-08-487-459-4	Sequence 4, Appl1
33	6	1.4	22	4	US-09-549-831-3	Sequence 3, Appl1
34	6	1.4	25	3	US-09-023-339-5	Sequence 5, Appl1
35	6	1.4	35	2	US-08-374-560-8	Sequence 8, Appl1
36	6	1.4	33	4	US-09-149-476-660	Sequence 660, App
37	6	1.4	34	1	US-07-915-247A-11	Sequence 11, Appl1
38	6	1.4	34	1	US-08-443-863-11	Sequence 11, Appl1
39	6	1.4	34	1	US-08-448-070-11	Sequence 11, Appl1
40	6	1.4	34	1	US-08-449-500-11	Sequence 11, Appl1
41	6	1.4	34	1	US-08-449-317A-11	Sequence 11, Appl1
42	6	1.4	34	2	US-08-477-022-11	Sequence 11, Appl1
43	6	1.4	34	2	US-08-449-447-11	Sequence 11, Appl1
44	6	1.4	34	2	US-08-184-328-11	Sequence 11, Appl1
45	6	1.4	34	2	US-08-521-097-11	Sequence 11, Appl1

ALIGNMENTS

RESULT 1
US-08-617-785-2
; Sequence 2, Application US/0861785E
; Patent No. 6228610
; GENERAL INFORMATION:
; APPLICANT: Flot, Peter J.
; APPLICANT: Kuhn, Rainer
; APPLICANT: Linder, Kristen
; APPLICANT: Putner, Irene
; APPLICANT: Knopfel, Thomas
; TITLE OF INVENTION: Human Metabotropic Glutamate Receptor Subtypes (HMR4,
; FILE REFERENCE: 4-19679/A/PCT
; CURRENT APPLICATION NUMBER: US/08/617,785E
; EARLIER FILING DATE: 1996-03-19
; EARLIER APPLICATION NUMBER: PCT/EP94/02991
; EARLIER FILING DATE: 1994-09-07
; EARLIER APPLICATION NUMBER: EPO 9416553.7
; EARLIER FILING DATE: 1994-08-19
; EARLIER APPLICATION NUMBER: EPO 93810663.0
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 2
; LENGTH: 912
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-617-785-2
Query Match 1.8%; Score 8; DB 4; Length 912;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 92 KAGEFDKI 99
DB 261 KAGEFDKI 268
RESULT 2
US-09-641-318-2
; Sequence 2, Application US/09641318
; Patent No. 6384205
; GENERAL INFORMATION:
; APPLICANT: BELAGAJE, RAMA M.
; TITLE OF INVENTION: EXCITATORY AMINO ACID RECEPTOR PROTEIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ELI LILLY AND COMPANY

THIS PAGE BLANK (USPTO)

XX Venter JC, Adams M, Li PWD, Myers EW;
 PI
 XX
 DR WPI: 2001-65660/75.
 DR N-PSDB; ABL04635.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 8388; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABB57737-ABB72072).
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1256 AA;

Query Match 1.8%; Score 8; DB 22; Length 1256;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 126 NLDKSLLEL 133
 |||||
 DB 1021 NLDKSLLEL 1028

Search completed: November 18, 2002, 12:03:25
 Job time : 45 secs

OS	Homosapiens	Location/Qualifiers
XX		
XX		
FT	Key	588..610
FT	Domain	/label= "TMD-I
FT		/note= "transmembrane domain I"
FT	Domain	625..645
FT		/label= "TMD-II
FT		/note= "transmembrane domain II"
FT	Domain	657..675
FT		/label= "TMD-III
FT		/note= "transmembrane domain III"
FT	Domain	699..720
FT		/label= "TMD-IV
FT		/note= "transmembrane domain IV"
FT	Domain	751..771
FT		/label= "TMD-V
FT		/note= "transmembrane domain V"
FT	Domain	786..807
FT		/label= "TMD-VI
FT		/note= "transmembrane domain VI"
FT	Domain	823..847
FT		/label= "TMD-VII
FT		/note= "transmembrane domain VII"
XX		
PN	W09522609-A2.	
XX		
XX	24-AUG-1995.	
PD		
PE	21-FEB-1995;	95WO-GB00356.
XX		
PR	01-AUG-1994;	94GB-0015532.
PR	21-FEB-1994;	94GB-0003285.
XX		
PA	(WELL) WELLCOME FOUND LTD.	
XX		
PI	Makoff AJ;	
XX		
DR	WPI; 1995-302715/39.	
DR	N-PSDB; AAT03888.	
XX		
PT	New isolated human metabotropic glutamate receptors - used for	
PT	detection, diagnosis and therapy of diseases associated with the	
PT	receptors, eg. stroke, epilepsy and Alzheimer's disease.	
XX		
PS	Claim 2; Page 40-43; 55pp; English.	
XX		
CC	mRNA from the human cerebellum was used to construct a cDNA	
CC	library. cDNA was amplified by PCR primers (AAT03896-97) based on rat	
CC	mgIuR4 sequences and with the primers given in AAT03898-99 to obtain	
CC	cDNA encoding human mgIuR4.	
XX		
SO	Sequence 912 AA;	
Qy	Query Match	1.8%; Score 8; DB 16; Length 912;
	Best Local Similarity	100.0%; Pred. No. 92;
Matches	8; Conservative	0; Mismatches
		0; Indels
		0; Gaps
Oy	92 KAGEFDKI 99	
Db	261 KAGEFDKI 268	
RESULT 14		
AAE23757		
ID	AAE23757 standard; Protein: 912 AA.	
XX		
AC	AAE23757;	
XX		
DT	10-SEP-2002 (first entry)	
XX		
DE	Human metabotropic glutamate (mgIuR4) receptor protein.	

XX	Human: metabotropic glutamate receptor; mGluR4; neurodegeneration;
KW	antipsychotic; anticonvulsant; analgesic; antidepressant; antiemetic.
XX	
OS	Homo sapiens.
XX	
PN	US6384205-B1.
XX	
PD	07-MAY-2002.
XX	
PF	18-AUG-2000; 2000US-0641318.
XX	
PR	12-MAR-1996; 96US-013189P.
PR	12-MAR-1997; 97US-0816178.
XX	
PA	(ELIL) LILLY & CO ELI.
XX	
PI	Belagaje RM, Wu S;
XX	
DR	WPI: 2002-442818/47.
DR	N-PSDB: AAD38024; AAD38025.
XX	
PT	New nucleic acid encoding human metabotropic glutamate receptor, useful
PT	e.g. in screening for specific agonists and antagonists for treating
PT	e.g. neurodegeneration
XX	
PS	Claim 1; Column 9-16; 35pp; English.
XX	
CC	The present invention relates to human metabotropic glutamate receptor
CC	(mGluR) proteins and polynucleotides encoding such proteins. mGluR4
CC	sequences of the invention are useful for treating acute and chronic
CC	neurodegeneration. They are also used as antipsychotic, anticonvulsant,
CC	analgesic, antidepressant and antiemetic agents. They are also useful
CC	for the diagnosis and/or treatment of conditions associated with an
CC	excess or deficiency of mGluR4. The present sequence is human mGluR4
CC	protein.
XX	
SQ	Sequence 912 AA;
XX	
QY	Query Match 1.8%; Score 8; DB 23; Length 912;
	Best Local Similarity 100.0%; Pred. No. 92;
Matches	8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	92 KAGEFDKI 99
DB	261 KAGEFDKI 268
XX	
RESULT 15	
ABB60532	
ID	ABB60532 standard; Protein: 1256 AA.
XX	
AC	ABB60532;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster polypeptide SEQ ID NO 8388.
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide;
XX	
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
XX	
PA	(PEKE) PE CORP NY.

CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX SQ Sequence 462 AA;

Query Match 1.8%; Score 8; DB 22; Length 462;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 VILVTLGA 22
|||||||
DB 438 VILVTLGA 445

RESULT 11

AA015101 ID AAO15101 standard; Protein; 909 AA.

XX AC AAO15101;

XX DT 22-AUG-2002 (first entry)

XX DE Human pH8SPmGluR4 chimeric protein.

XX DE Human: G-protein fusion receptor; extracellular domain;

KW transmembrane domain; intracellular domain; Car; mGluR; GABABR;

KM modulator identification.

XX OS Chimeric - Homo sapiens.

XX PN WO200229033-A2.

XX PD 11-APR-2002.

XX PF 03-OCT-2001; 2001WO-US31074.

XX PR 03-OCT-2000; 2000US-0679664.

XX PA (NPSF-) NPS PHARM INC.

XX PI Stormann T, Hammerland LG, Storjohann LL, Busby JG, Garrett JE;

XX PI Simin RT;

XX DR WPI; 2002-330170/36.

XX PT Novel G-protein fusion receptor, useful for identifying modulators of
XX Car, mGluR and GABABR, comprises G-protein joined to the intracellular
XX domain of the receptor -
XX PS Disclosure; Fig 16; 168pp; English.

XX CC The invention comprises G-protein fusion receptors - comprising
XX extracellular, transmembrane and intracellular domains similar to Car,
XX mGluR or GABAB receptor sequences. The G-protein fusion receptors of the
XX invention may also possess a linker joined to the carboxy terminus of the
XX intracellular domain, and a G-protein joined to the linker. The G-protein
XX fusion receptors of the invention are useful for identifying modulators
XX of Car, mGluR and GABABR for use in treating associated conditions. The
XX present amino acid sequence was used in the production of the invention.

XX SQ Sequence 909 AA;

Query Match 1.8%; Score 8; DB 23; Length 909;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 92 KAGEFDKI 99
|||||||
DB 258 KAGEFDKI 265

RESULT 12

AA072092 ID AAR72092 standard; Protein; 912 AA.

XX AC AAR72092;

XX DT 26-SEP-1995 (first entry)

XX DE Human mGluR4.

KW Human metabotropic glutamate receptor subtype 4; mGluR4; hmGluR4;

KM signal transducer.

XX OS Homo sapiens.

XX PN WO9508627-A.

XX PD 30-MAR-1995.

XX PF 07-SEP-1994; 94WO-EP02991.

XX PR 19-AUG-1994; 94GB-0016553.

XX PR 20-SEP-1993; 93EP-0810663.

XX PA (CIBA) CIBA GEIGY AG.

XX PI Flor PJ, Knoepfel T, Kuhn R, Lindauer K, Puettner I;

XX DR WPI; 1995-139596/18.

XX DR N-PSDB; AA089342.

XX PT Human metabotropic glutamate receptor sub-type(s), esp. 4, 6 and
XX 7 - also corresp. DNA and antibodies, useful for identifying
XX cpds. which modulate signal transduction activity

XX PS Claim 2; Page 44-48; 110pp; English.

XX CC Human metabotropic glutamate receptor subtype 4 (hmGluR4) cDNA

XX CC clones were isolated from a cerebellum cDNA library using a rat

XX CC mGluR4 probe. Clone cmR20 lacked the 5' end of the hmGluR4 gene.

XX CC PCR using human genomic or brain cDNA as template was used to

XX CC obtain a complete gene sequence (given in AA089342) encoding hmGluR4

XX CC (AAR72092). Recombinant hmGluR4 was produced in mammalian cells.

XX SQ Sequence 912 AA;

OY 92 KAGEFDKI 99
|||||||
DB 261 KAGEFDKI 268

RESULT 13

AA082658 ID AAR82658 standard; Protein; 912 AA.

XX AC AAR82658;

XX DT 20-DEC-1995 (first entry)

XX DE Human mGluR4.

KW Metabotropic glutamate receptor 4; mGluR4; stroke; epilepsy;

KW Alzheimer disease; detection; diagnosis; therapy.

CC Fragments of MTRP polynucleotides are useful as targets in microarrays.
CC MTRP DNA is also useful for generating hybridisation probes useful in
CC mapping genomic sequences and detecting differences in sequences among
CC normal, carrier and affected individuals. It is also used for
CC screening libraries of compounds in drug screening techniques.

XX Sequence 462 AA;

SO Query Match 1.8%; Score 8; DB 21; Length 462;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 VLVVTIGA 22
|||||||
Db 438 VLVVTIGA 445

RESULT 9
AAU14349
ID AAU14349 standard; Protein; 462 AA.

XX AAU14349;

XX 24-OCT-2001 (first entry)

XX Human novel protein #220.

XX Human: novel protein; Antianaemic; osteopathic; antiinflammatory;
KM immunomodulatory; cytostatic; neuroprotective; vulnerrary; nootropic;
KM anticonvulsant; antirheitic; cerebroprotective; antifungal; antiviral;
KM antibacterial; antiallergic; dermatological; haemostatic; antiashtmatic;
KM thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KM Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KM tissue regeneration; immune disorder.

XX Homo sapiens.

XX WO200155437-A2.

XX 02-AUG-2001.

XX 25-JAN-2001; 2001WO-US02623.

XX 25-JAN-2000; 2000US-0491404.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI: 2001-451939/48.

XX N-PSDB; AAS22654.

PT Isolated polypeptides useful for treating anti-inflammatory diseases,
XX nervous system disorders, and for regenerating bone and cartilage -

XX Example 4; Page 655-656; 894pp; English.

CC The invention relates to polynucleotides encoding novel human
CC proteins or their active domains. The polypeptides, polynucleotides and
CC antibodies raised against the polypeptides are used in a method of
CC treatment of a mammal and prevention of disorders caused by the aberrant
CC protein expression or activity. The polypeptides can be used as
CC molecular weight markers, food supplements, and in antibody production.
CC The polypeptides are used to identify compounds which bind to the
CC polypeptides. Polynucleotides of the invention are used as probes and
CC primers, for sequencing, for chromosome or gene mapping, in the
CC production of recombinant proteins, and in generating anti-sense DNA or
CC RNA and in gene therapy. Polypeptides of the invention can be used to
CC target drugs to a tumour, in assays to determine biological activity, to
CC raise antibodies/elicit an immune response, to determine quantitative
CC protein levels, as tissue markers, and to isolate receptors or ligands.
CC Polypeptides of the invention may also be useful in treating platelet
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,

CC ligament and/or nerve tissue, wound healing, treating burns, promoting
CC the proliferation, differentiation and survival of stem cells, as a
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC fungal infection or from autoimmunity, cancer, allergy, asthma,
CC graft-versus-host disease, eczema, haemophilia, thrombosis,
CC anti-inflammatory diseases, nervous system disorders, and infection.
CC The present sequence represents a protein of the invention.

XX Sequence 462 AA;

SO Query Match 1.8%; Score 8; DB 22; Length 462;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 VLVVTIGA 22
|||||||
Db 438 VLVVTIGA 445

RESULT 10
AAB93208
ID AAB93208 standard; Protein; 462 AA.

XX AAB93208;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:12180.

XX Human: primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI: 2001-318749/34.

XX Claim 8; SEQ ID 12180; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in

XX Novel human diagnostic protein #6777.
 DE
 KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR N-PSDB: AAS70973.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID No 37145; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 460 AA;
 Query Match 1.8%; Score 8; DB 22; Length 460;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 15 VLLVTLGA 22
 Db 436 VLLVTLGA 443
 RESULT 8
 AAY71066
 ID AAY71066 standard; Protein; 462 AA.
 XX
 AC AAY71066;
 XX
 DT 29-AUG-2000 (first entry)
 XX
 DE Human membrane transport protein, MTRP-11.

XX Human: membrane transport protein; MTRP-11; antiinflammatory; cytostatic;
 KW antithyroid; immunosuppressive; thymine; antidiabetic; noctropic;
 KW antidiabetic; neuroprotective; antidepressant; nephrotoxic; virucide;
 KW antihelminthic; protozoacide; antibacterial; neuroleptic; anti-gout;
 KW diagnosis; prevention; treatment; membrane transport disorder; epilepsy;
 KW Menkes disease; diabetes; Parkinson's disease; neurological disorder;
 KW Alzheimer's disease; depression; schizophrenia; immune disorder; allergy;
 KW inflammatory disorder; AIDS; Addison's disease; atherosclerosis; gout;
 KW Graves disease; Hashimoto's thyroiditis; microbial infection; cancer;
 KW cell proliferative disorder.
 XX
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT Modified-site 11
 FT /note= "phosphorylation site"
 FT Modified-site 37
 FT /note= "phosphorylation site"
 FT Modified-site 114
 FT /note= "phosphorylation site"
 FT Modified-site 174
 FT /note= "phosphorylation site"
 FT Modified-site 179
 FT /note= "phosphorylation site"
 FT Modified-site 420
 FT /note= "phosphorylation site"
 FT Region 58..104
 FT /note= "Aromatic amino acid permease signature sequence"
 FT Region 243..265
 FT /note= "Aromatic amino acid permease signature sequence"
 FT Region 375..394
 FT /note= "Aromatic amino acid permease signature sequence"
 PN WO200026245-A2.
 PD 11-MAY-2000.
 XX
 PF 04-NOV-1999; 99WO-US26048.
 XX
 PR 04-NOV-1998; 98US-0172255.
 PR 24-NOV-1998; 98US-0172252.
 PR 22-DEC-1998; 98US-0172214.
 PR 26-FEB-1999; 99US-0121896.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Hillman JL, Yue H, Tang YT, Lal P, Corley NC, Guegler KJ;
 PI Baughn MR, Azimzal Y, Lu DAM;
 XX
 DR WPI: 2000-365576/31.
 DR N-PSDB: AAD00610.
 XX
 PT Novel human membrane transport proteins useful for diagnosis,
 PT prevention and treatment of membrane transport disorders,
 PT immune/inflammatory disorders and cell proliferative disorders
 PT including cancer -
 XX
 PS Claim 1; Page 103-104; 136pp; English.
 XX
 CC The present sequence is a membrane transport protein,
 CC MTRP-11 from incyte clone 1413743 isolated from human BRAINOT12 cDNA
 CC library. MTRP-11 shows homology to Schistosomaccharomyces pombe
 CC transmembrane transporter and is expressed in nervous,
 CC haematopoietic/immune and reproductive tissues.
 CC The present sequence is useful in diagnosis, prevention and treatment
 CC of disorders related with increased or decreased expression of MTRP
 CC such as familial goitre, Menkes disease, diabetes, Parkinson's disease,
 CC neurological disorders such as Alzheimer's disease, depression, epilepsy,
 CC schizophrenia, immune/inflammatory disorders such as AIDS, Addison's
 CC disease, allergies, atherosclerosis, Graves disease, gout, Hashimoto's
 CC thyroiditis, viral, bacterial, fungal, parasitic, protozoal or
 CC helminthic infections and cell proliferative disorders such as cancer.

XX 13-FEB-2002 (first entry)
DE Novel human diagnostic protein #2016.
XX
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM Food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
XX N-PSDB; AAS66212.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID NO 32384; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 356 AA;
XX
XX Query Match 1.8%; Score 8; DB 22; Length 356;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 15 VLLVLTGA 22
DB 332 VLLVLTGA 339
XX
XX RESULT 6
AAU63605
ID AAU63605 standard; Protein; 420 AA.
XX
AC AAU63605;
XX
DT 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #24501.
DE
XX
KM SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA.
KM Inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
XX 02-JUN-2000; 2000US-208841P.
XX 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI: 2001-616774/71.
XX N-PSDB; AAS59634.
XX
XX Claim 3; SEQ ID NO 24800; 106pp; English.
XX
XX Sequences AAU3105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 420 AA;
XX
XX Query Match 1.8%; Score 8; DB 22; Length 420;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 8 RLGLSSV 15
DB 324 RLGLSSV 331
XX
XX RESULT 7
ABG06786
ID ABG06786 standard; Protein; 460 AA.
XX
AC ABG06786;
XX
DT 13-FEB-2002 (first entry)

degenerative disorder; osteoarthritis; neurodegenerative disorder;
 cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 hyperextension; hypothyroidism; cholesterol ester storage disease;
 immune deficiency; immune disorder; infectious disease;
 autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 myasthenia gravis.
 XX Homo sapiens.
 XX WO200192523-A2.
 XX 06-DEC-2001.
 XX 29-MAY-2001; 2001WO-US10836.
 XX 30-MAY-2000; 2000US-206132P.
 XX 29-AUG-2000; 2000US-228716P.
 XX (CURA-) CURAGEN CORP.
 XX Shinkets RA, Leach MD;
 DR WPI: 2002-106308/14.
 DR N-PSDB; ABN19148.
 XX Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and autoimmune disorders
 PS Disclosure; SEQ ID 6774; 1037pp; English.
 XX The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX

Sequence 78 AA:
 SQ
 Query Match 1.8%; Score 8; DB 23; Length 78;
 Best Local Similarity 100.0%; Pred. No. 9.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 418 DLKVS DLY 425
 |||||
 Db 2 DLKVS DLY 9

RESULT 4
 AAU60666
 ID AAU60666 standard; Protein: 236 AA.
 XX
 AC AAU60666;

27-FEB-2002 (first entry)
 XX Propionibacterium acnes immunogenic protein #21562.
 DE
 XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 XX dermatological; osteopathic; neuroprotectant.
 XX Propionibacterium acnes.
 XX WO200181581-A2.
 XX 01-NOV-2001.
 XX 20-APR-2001; 2001WO-US12865.
 XX 21-APR-2000; 2000US-199047P.
 XX 02-JUN-2000; 2000US-208841P.
 XX 07-JUL-2000; 2000US-216747P.
 XX (CORI-) CORIXA CORP.
 XX Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
 DR WPI: 2001-616774/71.
 DR N-PSDB; AAS59612.
 XX
 XX Example 1; SEQ ID No 21861; 1069pp; English.
 XX Sequences AAU9105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX

Sequence 236 AA:
 SQ
 Query Match 1.8%; Score 8; DB 22; Length 236;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 418 DLKVS DLY 425
 |||||
 Db 85 DLKVS DLY 92

RESULT 5
 ABG02025
 ID ABG02025 standard; Protein: 356 AA.
 XX
 AC ABG02025;

PT New isolated Helicobacter polynucleotides - used to develop products
 PT for the diagnosis, prevention and treatment of Helicobacter
 PT infections and gastrointestinal diseases
 XX
 PS Claim 8; Page 395-398; 2054pp; English.
 XX
 CC This sequence represents a Helicobacter pylori GPO protein of the
 CC invention. The polypeptides can be used for preventing or treating
 CC Helicobacter infections, and gastroduodenal diseases associated with
 CC these infections, including acute, chronic, and atrophic gastritis, and
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
 CC used for the production of antibodies. The products can also be used for
 CC detection and diagnosis.
 CC
 SQ Sequence 444 AA;
 Query Match 100.0%; Score 444; DB 19; Length 444;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 444; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKHFSVRLIGLSVLLVTIGASMAHQAQSYLPKHESVTLKNGLOVSVPLENKTGVIEVDY 60
 DB 1 MKHFSVRLIGLSVLLVTIGASMAHQAQSYLPKHESVTLKNGLOVSVPLENKTGVIEVDY 60
 QY 61 LYVGSNEMTSGSGIAHMLEHNFKSTKNLKAGEFDKIVKREGVSNASTSPDITRYFI 120
 DB 61 LYVGSNEMTSGSGIAHMLEHNFKSTKNLKAGEFDKIVKREGVSNASTSPDITRYFI 120
 QY 121 KTSQANLDSKLELFAETMGSNLKEDFELPEROVAEERWRDINSYIGMLYFFFTAY 180
 DB 121 KTSQANLDSKLELFAETMGSNLKEDFELPEROVAEERWRDINSYIGMLYFFFTAY 180
 QY 181 VYHPYHMTPIGFMDLQNTLKDIFKHFSLYQPKNAIVLVGVDNSQKFEELSKHFEES 240
 DB 181 VYHPYHMTPIGFMDLQNTLKDIFKHFSLYQPKNAIVLVGVDNSQKFEELSKHFEES 240
 QY 241 LKLNDEAIAITPYMKERKQAGARTAVVHKDGVHLEWVALYKVPFAFKHKOVALDALSR 300
 DB 241 LKLNDEAIAITPYMKERKQAGARTAVVHKDGVHLEWVALYKVPFAFKHKOVALDALSR 300
 QY 301 LGECKSSWLOSELVDKRLKLSQAFSHNMOLDESVEFLFINGNPNVKAALQKEIVALLE 360
 DB 301 LGECKSSWLOSELVDKRLKLSQAFSHNMOLDESVEFLFINGNPNVKAALQKEIVALLE 360
 QY 361 KLRKGEITQAELEKLNQKADFTISNLESSSDVAGLFADYLVQNDIGLTDYOROFDLK 420
 DB 361 KLRKGEITQAELEKLNQKADFTISNLESSSDVAGLFADYLVQNDIGLTDYOROFDLK 420
 QY 421 VSDLYRVANEFKDTOSTYVFLKP 444
 DB 421 VSDLYRVANEFKDTOSTYVFLKP 444
 RESULT 2
 ID AAB63147 standard; Protein; 42 AA.
 XX AAB63147;
 XX
 DT 26-MAR-2001 (first entry)
 XX
 DE Human secreted protein sequence encoded by gene 14 SEQ ID NO:73.
 XX
 KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
 KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotrophic;
 KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
 KW fungicide; ophthalmological; gene therapy; autoimmune disease; neoplasm;
 KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
 KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
 KW anglogenesis; nervous system disorder; Alzheimer's disease; infection;
 KW ocular disorder; corneal infection; wound healing; skin aging;
 KW food additive; preservative.
 XX

OS Homo sapiens.
 XX
 PN W0200061629-A1.
 XX
 PD 19-OCT-2000.
 XX
 PF 06-APR-2000; 2000MO-US09071.
 XX
 PR 09-APR-1999; 99US-0128694.
 XX
 PR 20-JAN-2000; 2000US-0176931.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX
 PI Ruben SM, Komatsoulis G;
 XX
 DR WPI, 2000-647420/62.
 DR N-PSDB; AAF22386.
 XX
 PT Isolated nucleic acid molecule encoding a human secreted protein is
 XX used in preventing, treating or ameliorating a medical condition -
 PS Claim 11; Page 465; 533pp; English.
 XX
 CC AAF22373 to AAF22421 encode the human secreted proteins given in AAB63134
 CC to AAB63182. AAB63183 represent more human secreted proteins
 CC and polypeptides homologous to them. Human secreted proteins have
 CC activities based on the tissues and cells the genes are expressed in.
 CC Examples of activities include: immunosuppressive; antiarthritic;
 CC antirheumatic; antiproliferative; cytostatic; cardiant; vasotrophic;
 CC cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
 CC fungicide; and ophthalmological. The polynucleotides and proteins can be
 CC used to prevent, treat or ameliorate a medical condition in e.g. humans,
 CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are
 CC also used in diagnosing a pathological condition or susceptibility to a
 CC pathological condition. Disorders which are diagnosed or treated include
 CC autoimmune diseases e.g. Rheumatoid arthritis, hyperproliferative
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular
 CC disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral
 CC ischaemia, anglogenesis, nervous system disorders e.g. Alzheimer's
 CC disease, infections caused by bacteria, viruses and fungi and ocular
 CC disorders e.g. corneal infection. The polypeptides can also be used to
 CC aid wound healing and epithelial cell proliferation, to prevent skin
 CC aging due to sunburn, to maintain organs before transplantation, for
 CC supporting cell culture of primary tissues, to regenerate tissues and in
 CC chemotaxis. The polypeptides can also be used as a food additive or
 CC preservative to increase or decrease storage capabilities. AAF22364 to
 CC AAF22372 and AAB63133 represent sequences used in the exemplification of
 CC the present invention.
 CC
 SQ Sequence 42 AA;
 Query Match 1.8%; Score 8; DB 21; Length 42;
 Best Local Similarity 100.0%; Pred. No. 5.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 15 VLVVLTGA 22
 DB 17 VLVVLTGA 24
 RESULT 3
 ID AABP03396 standard; Protein; 78 AA.
 XX AABP03396;
 XX
 DT 25-JUN-2002 (first entry)
 XX
 DE Human ORFX protein sequence SEQ ID NO:6774.
 XX
 KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KW

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 11:59:10 ; Search time 40 seconds
(without alignments)
1479.082 Million cell updates/sec

Title: US-09-881-752A-212
Perfect score: 444
Sequence: 1 MKHFSVKRLGLSSVLVLT.....VRVANEYKDTOSTTFELKP 444

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_101002.*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	444	100.0	444	19 AAW98306
2	8	1.8	42	21 AAB63147
3	8	1.8	78	23 AAB03396
4	8	1.8	236	22 AAB06066
5	8	1.8	356	22 AAB02025
6	8	1.8	420	22 AAB63605
7	8	1.8	460	22 AAB06786
8	8	1.8	462	21 AAB71066
9	8	1.8	462	22 AAB14349
10	8	1.8	462	22 AAB93208

11	8	1.8	909	23 AAO15101
12	8	1.8	912	16 AAR72092
13	8	1.8	912	16 AAR82658
14	8	1.8	912	23 AAE23757
15	8	1.8	1256	22 ABB60532
16	8	1.8	1422	23 AAO15102
17	7	1.6	11	16 AAR90269
18	7	1.6	11	20 AAY10769
19	7	1.6	11	23 AAE22482
20	7	1.6	63	21 AAB44937
21	7	1.6	63	21 AAG12428
22	7	1.6	63	21 AAB60725
23	7	1.6	71	22 AAB18945
24	7	1.6	81	21 AAG12495
25	7	1.6	103	21 AAG60133
26	7	1.6	111	21 AAG12494
27	7	1.6	117	22 AAB7411
28	7	1.6	119	22 AAB24301
29	7	1.6	124	21 AAG26740
30	7	1.6	134	22 AAB05640
31	7	1.6	136	22 ABB96064
32	7	1.6	136	22 AAB05372
33	7	1.6	141	18 AAB31859
34	7	1.6	141	22 AAO05860
35	7	1.6	142	22 AAB14281
36	7	1.6	143	23 AAB06537
37	7	1.6	147	18 AAB20468
38	7	1.6	164	21 AAB25275
39	7	1.6	165	23 AAB02336
40	7	1.6	166	22 AAB29981
41	7	1.6	166	23 AAE15840
42	7	1.6	173	21 AAG58958
43	7	1.6	187	22 AAB03928
44	7	1.6	187	23 ABB57411
45	7	1.6	188	21 AAB51647

ALIGNMENTS

RESULT 1	
AAW98306	AAW98306 standard; Protein: 444 AA.
ID	
AC	AAW98306;
DT	31-MAR-1999 (first entry)
XX	
DE	H. pylori GHP0 525 protein.
KW	GHP0 protein; Helicobacter infection; gastroduodenal disease; gastritis; peptic ulcer disease.
XX	
OS	Helicobacter pylori.
XX	
PN	W09843478-A1.
XX	
PD	08-OCT-1998.
XX	
PF	01-APR-1998; 98WO-US06371.
XX	
PR	29-JUL-1997; 97US-0902615.
PR	01-APR-1997; 97US-0833457.
PR	24-JUN-1997; 97US-0881227.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
PA	(IMMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX	
PI	Al-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;
XX	WPI; 1998-542293/46.
XX	
DR	N-PSDB; AAX14025.
XX	

Human p18SpmGluR4
Human mGluR4. Hom
Human mGluR4. Hom
Human metabotropic
Drosophila melanog
Human p18GluR4-Car
Ion-channel form in
peptide used to ma
Biologically activ
Human secreted pro
Zea mays protein f
Human secreted pro
Novel human diagno
Zea mays protein f
Zea mays protein f
Zea mays protein f
Human immune/hema
Human EST encoded
Zea mays protein f
Protonibacterium
Human testicular a
Human reproductive
Sunflower 2S album
Human polypeptide
Novel human diagno
Human ORFX protein
H. pylori secreted
Eucalyptus grandis
Human ORFX protein
Novel human secret
Helicobacter pylori
Zea mays protein f
Human gene 31 enco
Human secreted pro
Human secreted pro

```

Db 257 EKIDFNIPLINKWRLDSISEGTTIPSEISFENTIENTNTIASYKQELIQOITTRLNL 316
QY 304 GKSSMLOSEL---VDK-----KRLASQAFSHNMQLOD-----ESVEFLIAGNP 344
Db 317 RLOOW-EKETENGVDANSFRTHLGKETLOSIFS--LQLDIYOXSKTIDKLFAPIA----- 369
QY 345 NVKAERAL-OKEIVALLEKLEKKKEITO-AELDKLKINOKADEFISNIESSDVAGLFADYLV 402
Db 370 SIKQOGFTQNEL-----SGEIKRLQLQNEKQNLNIRS---GSLKTIADDLITSVANKOV 418
QY 403 QNDIQGLDIYORFL-DIKVSDIVRVANE 430
Db 419 VLSVNDREYELNKRFLSQITTLADLQRTLNQ 447

```

Search completed: November 18, 2002, 11:57:55
 Job time : 17 secs

RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kisser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Friesen T., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kisser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RA "Complete genome sequence of the model actinomycete Streptomyces
 RL coelicolor A3(2).";
 CC Nature 417:141-147(2002).
 CC -1- COFACTOR: REQUIRES Divalent cations for activity. binds zinc (by
 CC similarity).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR VAL-35 IS THE INITIATOR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL, AL031260; CAA20289.1; -
 CC DR MEROPS, M16; UPW; -
 CC DR InterPro: IPR001431; Peptidase_M16.
 CC DR Pfam: PF00675; Peptidase_M16; 1.
 CC DR PROSITE: PS00143; INSULINASE; 1.
 CC KW Hypothetical protein; Hydrolase; Metalloprotease; zinc;
 CC Complete proteome.
 CC FT METAL 79 79 ZINC (BY SIMILARITY).
 CC FT ACT SITE 82 82 BY SIMILARITY.
 CC FT METAL 83 83 ZINC (BY SIMILARITY).
 CC FT METAL 159 159 ZINC (BY SIMILARITY).
 CC SQ SEQUENCE 459 AA; 49671 MW; 8B9CFC166E11700F CRC64;

Query Match 11.3%; Score 256.5; DB 1; Length 459;
 Best Local Similarity 24.0%; Pred. No. 4e-09;
 Matches 93; Conservative 68; Mismatches 179; Indels 47; Gaps 8;

QY 37 TLKNGIQQVSVPLENTGVEVDVLYKYSRNETMGSGIAHMLEHLEINFKSTKNLAGEF 96
 DB 39 TLPGRLITETLPYSVRS-TEGIMAHVGSROETPALNATHLEHLEKGTKRSLDI 97
 QY 97 DKIVKRGVSNASTSFDTIRFIKTSQANLDSLELFAETMGSLNLEKEDFLPERQVA 156
 DB 98 SSALDAVGGEMNAFTAKECTCYARVLDPLAIDVCDMLGSLIOEDVDVERGAIL 157
 QY 157 EERRRMTDPSIGMLYFRFNTAYVYHPRHWPIDGFMDDIQNTLKDIFKHSLYQPN 216
 DB 158 EELIA-MTEDDPGDCCVHDLFAHTMFGDNLGRVLTVDVNAITADRIREFYKHHDPN 216
 QY 217 AIVLVGDVNSQKVELSKHFESEKLNLEKAIPTPYMKERKOD-----GARTAV 266
 DB 217 LVVAAGNDVHKVYQVAAAEKSGALDPPAAQPLAPAGRTYVAAGRVELIGKTTQ 276
 QY 267 VKDGVHLEWALGYKVPFKNHQ--VALDALSRLLGEGKSSWLOSELVDKKRLASQAF 324
 DB 277 AH-----VILG-MPLARTDERRWAMGVINTALGCMSSRLLEQVEYREKGLASYV 326
 QY 325 SHNQLODESVPFLTAGSPNKAELQKEIVALLEKLGELTQAEIDKIKINQADT 384
 DB 327 STTSGFADGGLFGVYAGCRPSOVHVL-----KICRELHVHAEHLTD-- 370
 QY 385 SNLESSSVAGLFPADLYQNDIOGLTD 411
 DB 371 -----DEIGRAVQGLQSGTVLGLSD 390

RESULT 12
 YMT1_CAEEL

ID YMT1_CAEEL STANDARD; PRT; 471 AA.
 AC P98080;
 DT 01-FEB-1996 (rel. 33, Created)
 DT 01-NOV-1997 (rel. 35, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE Hypothetical protein F56D2.1 in chromosome III.
 GN F56D2.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RC SEQUENCE FROM N.A.
 RA STRAIN=Bristol N2;
 RL Du Z.;
 RL Submitted (AUG-1994) to the EMBL/Genbank/DBJ databases.
 RP REVISIONS.
 RA Du Z.;
 RA Submitted (APR-1997) to the EMBL/Genbank/DBJ databases.
 RL -1- CAUTION: DOES NOT SEEM TO HAVE A PROTEASE ACTIVITY AS IT LACK THE
 CC ZINC-BINDING SITE.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL, U13644; AAB52679.1; -
 CC DR MEROPS, M16; UNB; -
 CC DR WormPeP, F56D2.1; CE11226.
 CC DR InterPro: IPR001431; Peptidase_M16.
 CC DR Pfam: PF00675; Peptidase_M16; 1.
 CC DR PROSITE: PS00143; INSULINASE; 1.
 CC KW Hypothetical protein.
 CC SQ SEQUENCE 471 AA; 51736 MW; 29DB386ED531A904 CRC64;

Query Match 11.2%; Score 253.5; DB 1; Length 471;
 Best Local Similarity 23.1%; Pred. No. 6.3e-09;
 Matches 100; Conservative 82; Mismatches 210; Indels 41; Gaps 13;

QY 10 LGISSVLATVLSAASHAQS-----YLKHESEVTLKNGIQQVSVPLENTGVEVD 59
 DB 5 LAVSSALRPALNSOVYNNASSAVSKVYLASAPQAEVTTLKNGFRVYTE--DNGSATATVG 62
 QY 60 VLYKVSERNETMGSGIAHMLEHLEINFKSTKNLAGEFDKIVKRGVSNASTSFDTIRYE 119
 DB 63 VWIETGSRRENKKNVNAFLERLHKGTGKRASAALESLNAGKLNLSFTERPDOTAVF 122
 QY 120 IRTSQANLDSLELFAETMGSLNLEKEDFLPERQVAEERRRMTDPSIGMLYFRFNTA 179
 DB 123 VQAGAQDVEKVVYDILADVLRNSKLEASTIDTERVNLKLELASDVOHQ--VLFDMHAAG 181
 QY 180 YVYHPRHWPIDGFMDDIQNTLKDIFKHSLYQPNALVLYVGDVNSQKVELSKHFE 239
 DB 182 FQGTPLALSVLGTSEISIPISAQOLKEWQEDHYRPVRLVSAVG--GVSNVSLADKPYG 240
 QY 240 SLKLNDEKAIPTPYMKERKODGAR--TAVYHK--DGVHLEWALGYKVPFKNHQVALD 295
 DB 241 DLSNEYPR-----KVPOVDGRTFGSEKRYRNDVNPMTAFAVEGYVHAKDLALO 293
 QY 296 ALSRLIGE-----GKSSWLOSELVDKKRLASQAFSHNM-----LQDESVE--LFTAGCN 343
 DB 294 IANQIFGQMDVYHATISRTAASRLYQK--IGHDHGVANLGHFNINRKDGLFLGIYFADAH 351
 QY 344 PNKAEALQKEIVALLEKLGKGTQAEIDKIKINQADJFNSLESSSVAGLFA-DIYV 402
 DB 352 DLNDTSGIMKSVAHKWKHLASA--ATEEEVAMAKNOFRMLYONLEINTQKAGFNKELLY 410

Db 211 LSRTLDYLDYLNRYKAPRMVLAAGVHHQLLDLAKHLSSRVYVEDAPGILTPC-- 268
QY 256 EPKODGARTAVVHK-DGVHLEWVALGYKVPFKHKDOVALDLSLGE-----GKSSWL 309
Db 269 --RTGSE--IRHDDALPLAHVAIVBGPMPANPDNTLYOVANAIITHYCTCGGVHL 324
QY 310 QSELVD---KKRLASQAFSHNMOLQ-----DESEVLEFIAGNPVKAEA 350
Db 325 SSPLASVAVANKLCSQFQTFNYSYSDTGLGAFHVCADMSIDDMVEF----- 371
QY 351 LQKRIVALLEKKGELTQALDLKLNQKADFISNLESSDVA-GLFADLYVQNDIGL 409
Db 372 LQGMWMLCTGSATSEVTRGK---NLRNA-LVSHLDGTPVCEIDGRSLDTYGRIRPL 426
QY 410 TDYORFELDKVSDLVRYANEFKD 434
Db 427 AEMSRIOEVDQAKMLRDLCSKTFYD 451
RESULT 10
MPPB_YEAST STANDARD: PRT: 462 AA.
AC PI0507:
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mitochondrial processing peptidase beta subunit, mitochondrial
DE Precursor (EC 3.4.24.64) (Beta-MPP) (PEP).
GN MASI OR MIF1 OR YLR163C OR I9632.10.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88312592; PubMed=3044780;
RA Witte C., Jensen R.E., Yaffe M.P., Schatz G.;
RT "MAS1, a gene essential for yeast mitochondrial assembly, encodes a
RT subunit of the mitochondrial processing protease.";
RL EMO J. 7:1439-1447(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Kucabalski A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kucabalski T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Mardis E., Meneses S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Rifkin L., Riles L., Talcott A., Trevasakis E., Vignati D.,
RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 21-32.
RX MEDLINE=91177897; PubMed=2007593;
RA Yang M., Gell Y., Oppinger W., Soda K., James P., Schatz G.;
RT "The MAS-1 encoded processing protease of yeast mitochondria.
RT Interaction of the purified enzyme with signal peptides and a
RT purified precursor protein.";
RL J. Biol. Chem. 266:6416-6423(1991).
CC -1- FUNCTION: THE MITOCHONDRIAL PROCESSING PROTEASE CLEAVES
CC PRECURSORS FROM MITOCHONDRIAL PROTEIN PRECURSORS.
CC CATALYTIC ACTIVITY: Release of N-terminal transit peptides from
CC precursor proteins imported into the mitochondrion, typically with
CC Arg in position P2.
CC -1- COFACTOR: REQUIRES DIVALENT CATIONS FOR ACTIVITY.
CC -1- SUBUNIT: HETERODIMER OF ALPHA AND BETA SUBUNITS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: X07649; CA30489.1; -;
CC DR EMBL: U51921; AAB67487.1; -;
CC DR PIR: S00552; S00552.
CC DR PIR: A38734; A38734.
CC DR MEROPS: M16.003; -;
CC DR SCD: S0004153; MASI.
CC DR InterPro: IPR001431; Peptidase_M16.
CC DR Pfam: PF00675; Peptidase_M16; 1.
CC DR PROSITE: PS00143; INSULINASE; 1.
CC KW Hydrolyase; Metalloprotease; Zinc; Mitochondrion; Transit peptide.
CC FT TRANSIT 1 20
CC FT CHAIN 21 462 MITOCHONDRIAL PROCESSING PEPTIDASE
CC FT METAL 70 70 ZINC (BY SIMILARITY).
CC FT ACT_SITE 73 73 BY SIMILARITY.
CC FT METAL 74 74 ZINC (BY SIMILARITY).
CC FT METAL 150 150 ZINC (BY SIMILARITY).
CC SQ SEQUENCE 462 AA; 51083 MW; 084CE63C495EDFC4 CRC64;
Query Match 11.4%; Score 259.5; DB 1; Length 462;
Best Local Similarity 23.7%; Pred. No. 2,66-09;
Matches 97; Conservative 78; Mismatches 194; Indels 41; Gaps 11;
QY 2 KHPSKRLIGLSSVLVTLGASMAHQSLPKHESVTLKNGLOVSVLPENKTGVIEDVL 61
Db 8 KFRNTRILLSTIS-----SQIPGTRISKLPNGLITAEVYIPN-TSSATVGI 53
QY 62 YKVGSRNETMCKSGIAHMLHLENFKSTKNLKAEPDKYKRRFGVSNASTSFDTIRPYK 121
Db 54 VDAGSRANVKNKNGAHLEHLEAFKGTIONRQOGIELEINIGSLNATYRENTVYAK 113
QY 122 TSQANLKSLELFAETMGSNLKKEDEFLPERQVAAEERRMRTDNPICMLYFRFNTAVY 181
Db 114 SLQEDIPAAVDILSDILTKSVLDNSAIEERDVITRESE-EVDKKYDEVFVDHLHEITYK 172
QY 182 YHPYHWPFGFMDIDQNTLMDIKKHSLLYQPKAIVLVGVNDSQVFELESKHF--- 238
Db 173 DQPLGRTILGPKTKSTRTDLADYITKNYKGDNRVLAAGADHEKLVQYAOQYFEHV 232
QY 239 --ESLKNLDEKAIPTP-----YMKPKODGARTAVVHKGVLHLEWVALGYKVPARKH 288
Db 233 PKSESPVPLSGPRGLPVFCGERIKENTLPTTHIALA-LEGV--SMSADYFVAL--- 286
QY 289 KDQVALDALSLRLEGKSSWLOSELV---DKRLASQAFSHNMOLQDESVF-LFIAGN 343
Db 287 ATOALVGNMMDRAIGTGTNS--PSPLAVAASONGSILANSYMSFSTSYADSGIMGWYIVDS 344
QY 344 PNVAKEALQKEIVALLLEKKGELTQALDLKLNQKADFISNLESSDVA 393
Db 345 NEHNVQTLVNELEKWKRIKSGISDAEVNRAKQALKALLSLDGGSTAI 394
RESULT 11
YLD8_STRCO STANDARD: PRT: 459 AA.
ID YLD8_STRCO
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical zinc protease SC05738 (EC 3.4.99.-).
GN SC05738 OR SC9A10.02.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomycetaceae; Streptomycos.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;

```

DR MIM; 191328;
DR InterPro: IPR001431; Peptidase_M16.
DR Pfam: PF00675; Peptidase_M16; 1.
DR PROSITE: PS00143; INSULINASE; FALSE_NEG.
KW Mitochondrion; Inner membrane; Electron transport; Respiratory chain;
KW Oxidoreductase; Transferrin; Polymorphism.
FT TRANSIT 1 34 MITOCHONDRION.
FT CHAIN 35 480 UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX
FT VARIANT 301 301 CORE PROTEIN I.
SQ SEQUENCE 480 AA; 52618 MW; AAB1528E0181679 CRC64;

Query Match
Best Local Similarity 11.7%; Score 265; DB 1; Length 480;
Matches 102; Conservative 87; Mismatches 194; Indels 54; Gaps 14;

QY 26 AQSILPKHESVTLKNGLOVSVPLENKGVEVDVLYKVGSRNETMKGSGIAHMLEHNF 85
DB 41 ALQFVPEYQVSLDNGLRVASSEOSOPT--CTVGWIDVGSREFETKKNAGYFLEHIAF 98
QY 86 KSRNKLAGEFDKIVKRFEGSVNASSTFDITRFIKTSQANLDSLELFAETMSLKE 145
DB 99 KGTNRPGSALEKEVESWGAHLNAYSTREHTAYIKALSKDLPRAVELLDIVQNCLEL 158
QY 146 DEFLEPEROVAEERRRMTDPSIGMLYFRFNTAYVYHPYHWPFGFMDIOWNTLKD 205
DB 159 SQLEKEDVILREMQ--ENDASMDVYVNYLHAATFOGTPLAQAVEGSEVNRKISRDLT 217
QY 206 KERSLYQPRNAIVLVGVNDSOKVELSKHESLK-NDEKAIPPPYKPEKODGART 264
DB 218 EYLSHTRKAPRMVLAAGVEHQQDLDAOKHLGIPWYAEADVPT--LTCPHFTSE- 274
QY 265 AVYHK-DGVLEWALGKYKPAFKHQVALDLSRLGE-----GKSSMLQSELY 314
DB 275 -IHRDLPALPAAVAVESPGWASPSVALQVANAIIIGHDCTYGGVHLSPFASGAV 333
QY 315 DKRLAS-QAFS-----HNNQLDESVEFLTAGNPVAKAALOKETVAL 358
DB 334 ANNLCSQFQFSTCYAETGLGAHFVCDRKIDD--MMTV-----LOGQMMRL 379
QY 359 LEKIKGEITQALDKIKINOKADEISNLESSDVA-GLEFADVLYVNDIGLDYQROFL 417
DB 380 CTSATESEVARGK---NIRRNA-IVSHLDGTPVCEDIGRSLLTYGRRIPLAWESRIA 434
QY 418 DLKVSQDLVRYANEFKD 434
DB 435 EVDASVVEITCSKYID 451

RESULT 9
UCRL_MOUSE STANDARD; PRT; 480 AA.
AC UCRL_MOUSE 09CZ13; 09CWL6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ubiqinol-cytochrome C reductase complex core protein I, mitochondrial precursor (EC 1.10.2.2).
GN UQCRC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawal J., Shingawa A., Shidata K., Yoshino M., Itoh M., Ishii Y.,
RA Atakawa T., Hara A., Fukunishi Y., Komno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

```

```

RA Fleischmann W., Gaasterland T., Glass C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriber L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Catinini F., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Yashinaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690(2001).
CC -1- FUNCTION: THIS IS A COMPONENT OF THE UBIQUINOL-CYTOCHROME C
CC REDUCTASE COMPLEX (COMPLEX I) OR CYTOCHROME B-C1 COMPLEX, WHICH
CC IS PART OF THE MITOCHONDRIAL RESPIRATORY CHAIN. THIS PROTEIN MAY
CC MEDIATE FORMATION OF THE COMPLEX BETWEEN CYTOCHROMES C AND C1 (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: QH(2) + 2 ferri-cytochrome c = Q + 2
CC ferri-cytochrome c
CC -1- SUBUNIT: BCL COMPLEX CONTAINS 11 SUBUNITS; 3 RESPIRATORY SUBUNITS
CC (CYTOCHROME B, CYTOCHROME C1, RIESKE PROTEIN), 2 CORE PROTEINS AND
CC 6 LOW-MOLECULAR WEIGHT PROTEINS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane (By
CC similarity).
CC -1- CAUTION: DOES NOT SEEM TO HAVE A PROTEASE ACTIVITY AS IT LACK THE
CC ZINC-BINDING SITE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AK013128; BAB28666.1; -
DR EMBL; AK010553; BAB27022.1; -.
DR HSSP; P31800; IBE3.
DR MGD; MG1:107876; Uqcrc1.
DR InterPro: IPR001431; Peptidase_M16.
DR Pfam: PF00675; Peptidase_M16; 1.
DR PROSITE: PS00143; INSULINASE; FALSE_NEG.
KW Mitochondrion; Inner membrane; Electron transport; Respiratory chain;
KW Oxidoreductase; Transferrin; Polymorphism.
FT TRANSIT 1 34 MITOCHONDRION (BY SIMILARITY).
FT CHAIN 35 480 UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX
FT VARIANT 301 301 CORE PROTEIN I.
FT CONFLICT 223 223 N->H (IN REF. 1; BAB27022).
FT CONFLICT 318 318 C->Y (IN REF. 1; BAB27022).
SQ SEQUENCE 480 AA; 52768 MW; 9C6D48DC9D5E429 CRC64;

Query Match
Best Local Similarity 11.5%; Score 261; DB 1; Length 480;
Matches 103; Conservative 87; Mismatches 197; Indels 58; Gaps 15;

QY 21 GASMAQSY--LPRHESVTLKNGLOVSVPLENKGVEVDVLYKVGSRNETMKGSGIAH 78
DB 34 GTATFAALQSVPEYQVSLDNGLRVASSEOSNHT--CTVGWIDVGSREFETKKNAGY 91
QY 79 MLEHLNFKSKRNKLAGEFDKIVKRFEGSVNASSTFDITRFIKTSQANLDSLELFAETM 138
DB 92 FLEHLNFKSKRNKRGNALEKEVESIGAHNLNAYSTREHTAYIKALSKDLPRAVELLDIV 151
QY 139 GSILNKDEFLEPEROVAEERRRMTDPSIGMLYFRFNTAYVYHPYHWPFGFMDIOWNTL 198
DB 152 QNSLSLEDSQLEKEDVILREMQ--ENDASMDVYVNYLHAATFOGTPLAQAVEGSEVNR 210
QY 199 WTLKDKIKFSLYQPRNAIVLVGVNDSOKVELSKHESLK-NDEKAIP--PPYMK 255

```


OY		9	LIGLSSVLLVTLTGASMAHSAOYLPKHESV----	TLKNGLOVVSVLEKNKGTVIEVDLYKYG	65
			: : : : : : : : :		
Dd		11	LITLVATVLLLP-GRLIATAA--LPDDEKITTGOLDNGLRIMTYPHAHPKQVMLIQIHFG		67
OY		66	SNETWGSKSIAHMLEHELNFKSTK-----NLAKGEFDKIYRFPGVSNASTSFDIRFYIK		121
			: : : : : : : : : : :		
Dd		68	SLQEDENELGVANFVEHHMFNFCTKTMPGNKVLETFSMOLRGGRVYNATYSIDETIYOVS		127
OY		122	---TSQANIDKSELELEPAETWGSLSLKEDFELEPVOVAEER----	WRDNS--PIGM	170
			: : : : : : : : : : :		
Dd		128	LPFTOKNOIQVMALFISESMSNATEEKKLEVDAREGVITEEMRAHDADAKWRTSQARP----		184
OY		171	LYFRFPNTAVYVHPHYHTPIGFMDLIQNPTLDIKKHSHLYQPKNALIVYGVDNSQKV		230
			: : : : : : : : : : :		
Dd		185	--FLANTFNILDR----EPGILMDVATVATPQLRQFYORWQOPNNMFIIVGDIDSKA		238
OY		231	FELSKHNEFSL---KNLIDKAIPTPYMKPPKODGARNAVHNDGVHLEVALGYKVPAEK		287
			: : : : : : : : : : :		
Dd		239	LALIDNLSKLRLANKAAERKWPT-----KAENHLRFNIINKENRVANGIALYLRIPMWQ		293
OY		288	HKD-----QVALDALSRLLGGEGKSSWIQS-ELYDKKRILASQAFSHMLODESVELFTA		340
			: : : : : : : : : : :		
Dd		294	VNDEOSFIRQAEBSMLVOLFNQRIDERSGSEL--KTISGTAARSVKTIAPDVQSIFEFRYN		351
OY		341	GGNPNWK--AEALQKEITYAL-----LEVLKAGEITQAEIDKLIKINOAKD----	FISWL	387
			: : : : : : : : : : :		
Dd		352	ARDNNMODANAALMELATVIDOHGFSAELELDVKSRTLRLWLNKANVDAQERDLRLMTLSRL		411
OY		388	ESSS	391	
Dd		412	ASSS	415	
<hr/>					
RESULT 8					
UCRL_HUMAN					
ID	UCRL_HUMAN	STANDARD:	PRT:	480 AA.	
AC	P31930:				
DT	01-JUL-1993 (Rel. 26, Created)				
DT	01-FEB-1995 (Rel. 31, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Ubiquitin1-cytochrome C reductase complex core protein I, mitochondrial precursor (EC 1.10.2.2).				
DE	precursor (EC 1.10.2.2).				
GN	UQCRC1.				
OS	Homo sapiens (Human) :				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
OX	NCBI_TaxId=9606;				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Placenta				

RX MEDLINE=94012661; PubMed=8407948.
 RA Hoffman G.G., Lee S., Christiano A.M., Chung-Honet L.C., Cheng W.,
 RA Katchman S., Uitto J., Greenspan D.S.;
 RT "Complete coding sequence, intron/exon organization, and chromosomal
 RT location of the gene for the core I protein of human ubiquinol-
 RT cytochrome c reductase.";
 RL J. Biol. Chem. 268:21113-21119(1993).
 RN [2]
 RN SEQUENCE FROM N.A., AND VARIANT ASN-301.
 RC TISSUE=Fibroblast;
 RX MEDLINE=9434830; PubMed=8069229;
 RA Islam M.M., Tanaka M., Suzuki H., Torii K., Hattori N., Ozawa T.;
 RT "A complete cDNA sequence for core I protein subunit of human
 RT ubiquinol-cytochrome c reductase.";
 RL Biochem. Mol. Biol. Int. 32:797-805(1994).
 RN [3]
 RN ERRATUM.
 RX MEDLINE=95038519; PubMed=7951059;
 RA Islam M.M., Tanaka M., Suzuki H., Torii K., Hattori N., Ozawa T.;
 RL Biochem. Mol. Biol. Int. 33:410-410(1994).
 RN [4]
 RN ERRATUM.
 RX MEDLINE=95072552; PubMed=7981668;
 RA Islam M.M., Tanaka M., Suzuki H., Torii K., Hattori N., Ozawa T.;
 RL Biochem. Mol. Biol. Int. 33:815-815(1994).
 RN [5]
 RN SEQUENCE OF 35-45.
 RP TISSUE=Liver;
 RX MEDLINE=94147969; PubMed=8313870;
 RA Hughes G.J., Frutiger S., Paquet N., Pasquali C., Sanchez J.-C.,
 RA Tissot J.-D., Baltrouch A., Appel R.D., Hochstrasser D.F.;
 RT "Human liver protein map: update 1993.";
 RL Electrophoresis 14:1216-1222(1993).
 RN [6]
 RN VARIANT ASN-301.
 RP PubMed=10453733;
 RX Valnot I., Kassiss J., Chretien D., de Lonlay P., Parfalt B.,
 RA Munnich A., Kachaner J., Rustin P., Roetig A.;
 RT "A mitochondrial cytochrome b mutation but no mutations of nuclearly
 RT encoded subunits in ubiquinol cytochrome c reductase (complex III)
 RT deficiency.";
 RL Hum. Genet. 104:460-466(1999).
 CC -1 FUNCTION: THIS IS A COMPONENT OF THE UBIQUINOL-CYTOCHROME C
 CC REDUCTASE COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH
 CC IS PART OF THE MITOCHONDRIAL RESPIRATORY CHAIN. THIS PROTEIN MAY
 CC MEDIANE FORMATION OF THE COMPLEX BETWEEN CYTOCHROMES C AND C1.
 CC -1 CATALYTIC ACTIVITY: QH(2) + 2 ferricytochrome c = Q + 2
 CC ferriocytochrome c.
 CC -1 SUBUNIT: B-C1 COMPLEX CONTAINS 11 SUBUNITS; 3 RESPIRATORY SUBUNITS
 CC (CYTOCHROME B, CYTOCHROME C1, RIESKE PROTEIN), 2 CORE PROTEINS AND
 CC 6 LOW-MOLECULAR WEIGHT PROTEINS.
 CC -1 SUBCELLULAR LOCATION: Mitochondrial inner membrane.
 CC -1 CAUTION: DOES NOT SEEM TO HAVE A PROTEASE ACTIVITY AS IT LACK THE
 CC ZINC-BINDING SITE.
 CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L16842; AAA20046.1; -;
 DR EMBL; D26485; BAA05495.1; -;
 DR HSSP; P31800; 1BB3.
 DR MEROPS; M16.973; -;
 DR SWISS-2DPAGE; P31930; HUMAN.
 DR PHCI-2DPAGE; P31930; -;
 DR HSC-2DPAGE; P31930; HUMAN.
 DR Stena-2DPAGE; P31930; -;
 DR Genew; HGNC:12585; UQCRC1.

OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_Taxid=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwim M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- CORRECTOR: REQUIRES DIVALENT CATIONS FOR ACTIVITY. BINDS ZINC (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AL008967; CA15577.1; -;
DR EMBL: AE007111; AK47171.1; -;
DR MEROPS: M16; UPW; -;
DR TIGR: MT2852; -;
DR Tuberculosis; Rv2782c; -;
DR InterPro: IPR001431; Peptidase_M16.
DR Pfam: PF00675; Peptidase_M16; 1.
DR PROSITE: PS00143; INSULINASE; 1.
KW Hypothetical protein; Hydrolase; Metalloprotease; zinc;
KW Complete proteome.
FT METAL 59 59 ZINC (BY SIMILARITY).
FT ACT_SITE 62 62 BY SIMILARITY.
FT METAL 63 63 ZINC (BY SIMILARITY).
FT METAL 139 139 ZINC (BY SIMILARITY).
FT CONFLICT 196 196 R -> W (IN REF. 2).
SQ SEQUENCE 438 AA; 47072 MW; C16707238706DE71 CRC64;

Query Match 12.5%; Score 283.5; DB 1; Length 438;
Best Local Similarity 27.5%; Pred. No. 7.5e-11;
Matches 111; Conservative 66; Mismatches 201; Indels 25; Gaps 9;

OY 37 TLKNGLOVVS--VPLENKTVIEVDLYKKYSHNEMTKSGIAHMLEHLMFKSTKNLKAG 94
DB 19 TLGGGLRVYTEFLPAYHSA---SVGVWVGVSDEGATYVGAHAHLEHLFKSTPTPSAV 75

OY 95 EFDKIVKREGVSNASTSPDITRYFIKTSQANDKSLSEFAETMGSLNKEDEFLPEROV 154
DB 76 DIQADMDAVGELNAFLAKHTCYVAHVGLSDLPVANDVADVVLNRCADADVEVERDV 135

OY 155 VAEERMRKDNPSIGMLYFFFTAVYHHPYHTPIGEMDDIONWTLKIDIKKFKHSLYOP 214
DB 136 VLEEIAMR-DDDEPDLADWFLAALFGDHPVGRPVIGSAQSVSMTRAQLSFHLRRYTP 194

OY 215 KNAIVLVGDVNSQKVFELSKHFEESLKNLDEKAIPTPYMKPEKODGA-RTAVVHKDGVH 273
DB 195 ERNVVAAGVNDHDLGVALVREHFGSRLVGRNPV-APRKGRGVNGSPRLTLVSDAQ 253

OY 274 LEVWALGYKVPAPKHKQVADLALSLRLGEGSSWISQSELDVKRLASQAFSHNMLQDE 333
DB 254 TH-VSIGIRTPGCGWHRRNALSVLHTALGGLSRFLQEVRETRGLAVSVS----- 304

OY 334 SVYFLTAGGNPNPKAALQKEI-----VALLEKLKKEIQAELDKIKNQKADFIEN 386
DB 305 ADLFLPDSGLSVYAACLPERFADVNRVTADVLESVARQGITAEACGIAGKSLRGGLVIG 364

OY 387 LE-SSSDVAGLFADYLVQNDIQGLTDPYOFDLTKVSDLVRYA 428
DB 365 LEQSSSRMSRLGSELSNLYCKRHSREHTLQIQEVYVEEVNAYV 407

RESULT 7
POOL.ECOLI
ID POOL.ECOLI STANDARD: PRT: 931 AA.
AC P31828; P31829; P76132; P78158;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable zinc protease p9ql (EC 3.4.99.-).
GN POOL OR B1494.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Turlin E., Gaesser F., Biville F.;
RT "Sequence and functional analysis of an Escherichia coli DNA fragment
RT able to complement p9ql and p9ql from Methylobacterium organophilum.";
RL Submitted (MAY-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Samei G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takekoshi K., Takeuchi Y., Mada C., Yamamoto Y., Horiuchi T.,
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
CC -1- SIMILARITY: STRONG, TO H.INFLUENZAE H11368.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 651 THAT PRODUCES TWO SEPARATE ORFS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----

DB 302 TRGLAVSVSALDIIFADSGALSVYAACLPGRFADVMQ-VISEVLASVAGDGITEACRIA 360
 QY 376 KINOKADFIENLE-SSSDVAGLFADYLVQNDIOGLTFYQOFDLTKVSDLYRVANE 430
 DB 361 KGSIRGSIITIGLEDSNWSMSRLGRSELNTRYKRGIEHTLQOIDEVYEQVALAHQ 416

RESULT 5
 UCRL BOVIN STANDARD; PRT; 480 AA.

AC P31800;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Ubiquinol-cytochrome C reductase complex core protein I, mitochondrial precursor (EC 1.10.2.2).
 GN UOCRC1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
 OC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=91293112; PubMed=1712295;
 RA Gencic S., Schaeffer H., von Jagow G.;
 RT "Core I protein of bovine ubiquinol-cytochrome-c reductase: an additional member of the mitochondrial-protein-processing family. Cloning of bovine core I and core II cDNAs and primary structure of the proteins.";
 RT Eur. J. Biochem. 199:123-131(1991).
 RN [2]
 RP REVISIONS TO C-TERMINUS.
 RX MEDLINE=96130702; PubMed=8592474;
 RA Schaeffer H., Brandt U., Gencic S., von Jagow G.;
 RT "Ubiquinol-cytochrome-c reductase from human and bovine mitochondria.";
 RT Meth. Enzymol. 260:82-96(1995).
 RN [3]
 RP SEQUENCE OF 35-52.
 RX MEDLINE=89062436; PubMed=2848575;
 RA Gonzalez-Halphen D., Lindorfer M.A., Capaldi R.M.;
 RT "Subunit arrangement in beef heart complex III.";
 RT Biochemistry 27:7021-7031(1988).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
 RX MEDLINE=9739328; PubMed=9204897;
 RA Xia D., Yu C.A., Kim H., Xia J.Z., Kachurin A.M., Zhang L., Yu L., Deisenhofer J.;
 RT "Crystal structure of the cytochrome bcl complex from bovine heart mitochondria.";
 RT Science 277:60-66(1997).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
 RX MEDLINE=98316377; PubMed=9651245;
 RA Iwata S., Lee J.W., Okada K., Lee J.K., Iwata M., Rasmussen B., Link T.A., Ramaswamy S., Jap B.K.;
 RT "Complete structure of the 11-subunit bovine mitochondrial cytochrome bcl complex.";
 RT Science 281:64-71(1998).
 RL -1- FUNCTION: THIS IS A COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS PART OF THE MITOCHONDRIAL RESPIRATORY CHAIN. THIS PROTEIN MAY MEDiate FORMATION OF THE COMPLEX BETWEEN CYTOCHROMES C AND C1.
 CC -1- CATALYTIC ACTIVITY: OH(2) + 2 ferricytochrome c = O + 2 ferriocytocrome c.
 CC -1- SUBUNIT: BCL COMPLEX CONTAINS 11 SUBUNITS; 3 RESPIRATORY SUBUNITS (CYTOCHROME B, CYTOCHROME C1, RIESKE PROTEIN), 2 CORE PROTEINS AND 6 LOW-MOLECULAR WEIGHT PROTEINS.
 CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane; matrix side.
 CC -1- CAUTION: DOES NOT SEEM TO HAVE A PROTEASE ACTIVITY AS IT LACK THE ZINC-BINDING SITE.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X59692; CAA42213.1; -.
 DR PIR: S16220; ZPEOCI.
 DR PDB: 1QCR; 14-OCT-98.
 DR PDB: 1BGY; 23-FEB-99.
 DR PDB: 1BE3; 16-FEB-99.
 DR MEROPS: M16.973; -.
 DR InterPro: IPR001431; Peptidase_M16.
 DR Pfam: PF00675; Peptidase_M16; 1.
 DR PROSITE: PS00143; INSULINASE; FALSE_NEG.
 KW Mitochondrion; Inner membrane; Electron transport; Respiratory chain;
 KW Oxidoreductase; Transit peptide; 3D-structure.
 FT TRANSIT 1 34 MITOCHONDRION
 FT CHAIN 35 480 UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX CORE PROTEIN I.
 FT SQ SEQUENCE 480 AA; 52735 MW; 4E54DD28C1905392 CRC64;

Query Match 13.1%; Score 297; DB 1; Length 480;
 Best local Similarity 23.2%; Pred. No. 1,2e-11;
 Matches 103; Conservative 96; Mismatches 189; Indels 56; Gaps 13;

QY 21 GASMAQSY--LPRHESVTLKGLQVSVPLENTGVLVDVLYKVSRNFTMGSGIAH 78
 DB 34 GTATVQAQLQSVPEFYQSQDNGLRVASEQSSQPT--CTGVGMWDAGSRYSEKKNAGY 91
 QY 79 MLEHLNKRSTKNLAGEEDKTVKRPFGVSNASTSFDIRYIKTSQALDLSLLEPAETM 138
 DB 92 FVEHLARFGTKNRPGMALEKEVESMAHLNASTREHAYVYIKALSKDLPRAVELADIV 151
 QY 139 GSLNKEDEFLPEROVAAEERRMRTDNPGLMFLRFENFTAVYVHPYHPTIGFMDION 198
 DB 152 ONCSLEDSQIKERDVIILQIQ-ENDSIMRDVENVYLIHATFOGTPLAQSVGESENVK 210
 QY 199 WTLDIRKFNHSLYQPKNAIVLVYGVNSQKVFELSKKHFEELSLK-NLDEKAIPTPYKEP 257
 DB 211 LSRADLFETLSRHYKARPVLAAGAEGLEHRLDLAQKHFGSGTYDEDAVPT--LSPC 268
 QY 258 KQDGAFTAVYK-DCVHLEWVALGKYVPAFKHKQVALDALSLRLGE-----GRSSWLQS 311
 DB 269 RFTGSQ--ICHRDEGLPLAHVAIVEGPGMAHPDVAIQVANAIIIGHYDCTYGGGAHLSS 326
 QY 312 EL-----VPKKRLASQAFSHNMLODESVFLFTAGGNPNVKAEL 351
 DB 327 PLASIAATNKLCSQSFQFTNICYADTGLGAHFFVCDHNSID--MMFY-----L 372
 QY 352 QKEIVALLEKLKKEITQAEELDKLNOKADFIENLESSSDVA-GLFADYLVQNDIOGLT 410
 DB 373 QGQNMRLCT-----SAFESEVLRGKNNLRNALVSHLDGTTPVCEDIGRSLTYGRRIPLA 427
 QY 411 DYQROFLDKVSDLYRVANEYFKD 434
 DB 428 EMESRIAEVDARVAVREVCSKYFYD 451

RESULT 6
 YR82_MYCTU STANDARD; PRT; 438 AA.
 AC O33324;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical zinc protease RV2782c (EC 3.4.99.-).
 GN RV2782C OR MT2852 OR MTV002.47C.
 OS Mycobacterium tuberculosis.

CC -1- COFACTOR: REQUIRES DIVALENT CATIONS FOR ACTIVITY. BINDS ZINC (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D27560; AAA3485.1; -
DR EMBL: L08471; AAA22379.1; -
DR EMBL: Z99112; CAB13544.1; -
DR PIR: S34595; S34595.
DR PIR: B46665; B46665.
DR MEROPS: M16.UPB; -
DR Subtilist; BG10779; ymxg.
DR InterPro: IPR001431; Peptidase_M16.
DR Pfam: PF00675; Peptidase_M16.1.
DR PROSITE: PS00143; INSULINASE; 1.
KW Hypothetical protein; Hydrolase; Metalloprotease; Zinc;
KW Complete proteome.
FT METAL 46 46 ZINC (BY SIMILARITY).
FT ACT_SITE 49 49 BY SIMILARITY.
FT METAL 50 50 ZINC (BY SIMILARITY).
FT METAL 126 126 ZINC (BY SIMILARITY).
FT CONFLICT 270 270 FQ -> LE (IN REF. 2).
FT CONFLICT 360 360 E -> Q (IN REF. 2).
SQ SEQUENCE 409 AA; 45963 MW; EA9BCF0EBD57563 CRC64;

Query Match 13.2%; Score 300; DB 1; Length 409;
Best Local Similarity 26.7%; Pred. No. 6.3e-12;
Matches 97; Conservative 70; Mismatches 170; Indels 26; Gaps 8;
QY 37 TLKNGLOVSVPLENKTGV--IEVDLYKYSRNETGKSGIAHMLEHLPKSTKNLAKG 94
D 6 TCNGVRIV---LENNPTVRSVAIGWIGSGHETPEINGISHLEHMFPGKTSKRSAR 62
QY 95 EFKIKYKRGCVSNASTSPDITRYFKTSQANDLSLEFAELMGSLNLEKEDEFLPEROV 154
D 63 ELASESDRIGGQVNAFTSKETCYAKVLDENHANYALDVLADMFESTFDENEIKKEKV 122
QY 155 VAEERWRDNDNSPIGLMFFEFNTAVYHYHPTPIGEMDIONMTLKDIKKPHSLYOP 214
D 123 VIEIIMYED-APDDIVHDLISKATYGNHSLGPILGTEETLASFNGDSIRQYMHDTTP 181
QY 215 KNAIVLVGDVNSQKFELESKHFEESLK-----NLDEKAIPTPYMKPEKQDGAATVAH 268
D 182 DRVVISVAGNISDSFIKDV-EKWFSGYEAKGKATGLEPPEFHTEKLRKE-----TEQAH 236
QY 269 KDGVHLEWVALGYKVPFAFKKDOVALDALSRLLGEGKSSWLGSELVDKRLASQAFSHM 328
D 237 -----LCIGFKGLGEVGERIYDLIVNNVGGSSRSRLFDVREDKGLAVSYSHS 288
QY 329 QIODESVLEFIAGCNPNVKAALOKKEIVALELKKGETQALDLKINOKKADPISNLE 368
D 289 SYED-SGMLITVGGTANDLOLSEITIOETLATLTKRGITSKLENSKEOMKSLMISLE 347
QY 389 SSS 391
D 348 STN 350

RESULT 4
YR82 MYCLE STANDARD; PRT; 445 AA.
AC 032965; 09C67;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical zinc protease ML0855 (EC 3.4.99.-).

GN ML0855 OR MLCB22.26C.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Fellingworth J., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Janczewska C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rastandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RA "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -1- COFACTOR: REQUIRES DIVALENT CATIONS FOR ACTIVITY. BINDS ZINC (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z98741; CAB1391.1; -
DR EMBL: AL583920; CAC31236.1; ALT_INIT.
DR Lepidoma; ML0855; -
DR MEROPS: M16.UPB; -
DR InterPro: IPR001431; Peptidase_M16.
DR Pfam: PF00675; Peptidase_M16.1.
DR PROSITE: PS00143; INSULINASE; 1.
KW Hypothetical protein; Hydrolase; Metalloprotease; Zinc;
KW Complete proteome.
FT METAL 66 66 ZINC (BY SIMILARITY).
FT ACT_SITE 69 69 BY SIMILARITY.
FT METAL 70 70 ZINC (BY SIMILARITY).
FT METAL 146 146 ZINC (BY SIMILARITY).
SQ SEQUENCE 445 AA; 48155 MW; 9AD8045F31D3D8 CRC64;

Query Match 13.2%; Score 299; DB 1; Length 445;
Best Local Similarity 27.6%; Pred. No. 8.2e-12;
Matches 115; Conservative 67; Mismatches 222; Indels 12; Gaps 7;
QY 21 GASMAHOSYLPKHESV---TLKNGLOVSVPLENKTGV--IEVDLYKYSRNETGKSG 75
D 7 GAEGKAEEKKARAGACRTLPGLRLVYT---EHLPAVVSASVGVWVGSREGATVAG 63
QY 76 IAHMLEHLPKSTKNLAKGFDKIVKRFGVSAVNSPDTREFTITSQANDKSLELA 135
D 64 AAHFLHLEKSTSTRTAMDIAQIDAVALGELAAFTAKEHTCYAHVLDDELAADLVA 123
QY 136 ETMGSLNLEKEDEFLPEROVAEERWRDNDNSPIGLMFFEFNTAVYHYHPTPIGEMD 195
D 124 DVLVNGRCANDVLEDRDYLLEIAMR-DDDPEDALGDMFLAALFDGHDYGRVITGMS 182
QY 196 IQNWLTKDKIKKFSLSLYOPKNAIVLVGDVNSQKFELESKHFEESLKNDKAIPTPYMK 255
D 183 VSAMTRTQHSFHYRRYTERRMVVAAGNVHDHDEVALVEHSGRLIRGOSAP-PRKS 241
QY 256 EPKQDGAATVAHKKDGVHLEWVALGYKVPFAFKKDOVALDALSRLLGEGKSSWLGSELVD 315
D 242 TGRINGPALLTIGKRAEDQTHVLLGVTRPGRSWEHMAVSLHTALAGGSLSLFOEIRE 301
QY 316 KRLASQAFSHNMLODESVEFLFIAGCNPNVKAALOKKEIVALELKKGETQALDKL 375

QY 68 NEMTSGIAHMLEHNFSTKNKAGCEPDKYKRGCVSNASTSDITRTYIKTSOANL 127
DB 121 DEPPGSGSLHFLHEHMFSTKHPSEFSKAKIAEIGGEENAFGSDYTAHVOTVPESTL 180
QY 128 DKSLLEAFETMGSLNKEDEFLPEROVAAEERMRRTDNSDGMXYFFFTAVYVHYHM 187
DB 181 RTMEEFPADMRHVLDAVIVPERDYLTERMRVENDPEQLLEEMQATLYONHRYRI 240
QY 188 TPIGFMDIONWTLKDKKHFSLYOPKNAIVLVGDVNSQKVELSKHFESEL-KMIDE 246
DB 241 PTIGMHEMQLNREDAKRYDRYAPNNAIIVAGDVDAGRVQADDEFGLTPRPDPL 300
QY 247 KAIPTPYKPEKODGATAVVHKDGYHL-----EWALGKVPKPKHKKDVALDALSLRL 301
DB 301 PAHVPR--QPEQONTKRIVALTDPRTVTPSFQKSWTTSYGTAA--EQGEAEALDILSEIL 356
QY 302 GEGKSSMLQSELYDKKRLAQAQASH-NMOQODESVFLFINGNPNVNAEALQKEIVALLE 360
DB 357 GGGTSTRTIYQELVAVKQIASSGGAYFNGRSLDPSSFTVFGSPRGEAKIEEVEDAIDAEIR 416
QY 361 KLKKEGTTQAEIDKIKINQKADFTSNLESSSDVAGLEFADYLVQND-IQGLTDYQROPLDL 419
DB 417 KIEFGTIDVELKAKNRFRSIIIFARDSSQSMAGIYGALATGDTAHVDYEWAPLRIKAV 476
QY 420 KVSDDLAVANEFKDTQSTTVFLKP 444
DB 477 KAAEVQAAARKYLSPDRSVAGYLLP 501
RESULT 2
Y219_RICPR STANDARD; PRT; 412 AA.
AC 005945;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical zinc protease Rp219 (EC 3.4.99.-).
GN Rp219.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=782;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=97419517; PubMed=9274032;
RA "Genomic rearrangements during evolution of the obligate
RT intracellular parasite Rickettsia prowazekii as inferred from an
RL analysis of 52015 bp nucleotide sequence.";
RN Microbiology 143:2783-2795(1997).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,
RA Scheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria";
RL Nature 396:133-140(1998).
CC -1- COFACTOR: REQUIRES Divalent cations for activity. binds zinc (by
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).-----

DR EMBL: Y11780; CAU72467.1; -.
DR EMBL: AJ235270; CAU14682.1; -.
DR MEROPS; M16.0PB; -.
DR InterPro: IPR001431; Peptidase_M16.
DR Pfam: PF00675; Peptidase_M16.1.
DR PROSITE: PS00143; INSULINASE; 1.
KW Hypothetical protein; Hydrolase; Metalloprotease; zinc;
KW Complete proteome.
FT METAL 49 49 ZINC (BY SIMILARITY).
FT ACT SITE 52 52 BY SIMILARITY.
FT METAL 53 53 ZINC (BY SIMILARITY).
FT METAL 129 129 ZINC (BY SIMILARITY).
SQ SEQUENCE 412 AA; 46611 MW; 29656AEB381031AE CRC64;
Query Match 15.0%; Score 340.5; DB 1; Length 412;
Best local similarity 25.8%; Pred. No. 1.8e-14;
Matches 108; Conservative 88; Mismatches 183; Indels 39; Gaps 15;
QY 38 LKNGLOVY--SVPLENTGVIEVDVLYKGSRNEMTCKSGIAHMLEHNFSTKNK-- 92
DB 10 LKNGLTITLYNMPYVHS---VAIHLAKVGARYENEDEGISHFLHEMAFGKTRPTAQ 66
QY 93 -AGFEDKIVKRGVSNASTSFDITRTYIKTSOANLKSLELFAETMGSLNKEDEFLPE 151
DB 67 IAEFFDSI---GGYFNAVYGHENTVYVYARVLSNCHKANLTIADIIQNSIFADEIATKE 122
QY 152 ROVVAEERMRRTDPSIGMLYFRFPNTAYVYHPYHPTPIGFMDDIQWTLKDKKHFSLY 211
DB 123 YQIMQELAHHDN-PDDLIYETFTYNTVYKQPLGKSLIGTKLYFTFHEHNFNFQKH 181
QY 212 YQPKNAIVLVGDVNSQKVELSKHFESEK-NDEKAIPTPYKPEKODGARTAVVHKD 270
DB 182 YNAENLYSLAGNIENHKIYVIAEELFASLQGVKSSFIPAKYI-----GKG-GFTIHE 234
QY 271 GVHLEWALGVKVPKPKHKKQV-ALDALSLRLGKGSMSQSELYDKKRLAQAQASHNMQ 329
DB 235 -LEQTSVLGSECTSYINLQYQYLLSIIFGGMSGRSLFQSTIREKLGAYVVSNSA 293
QY 330 LQDSVFLFTAGNPKNKAEAL-----QKEIVALEKIKKEITQAEIDKIKINQKADFTS 385
DB 294 YFDSGVTTIYASTAHN-KLELIYREITNEIKITETVSTEEITIAKM-QLRSLNQMQEQ 351
QY 386 NLESSDVA---GLFADYLVQNDIQGLTDYQROPLDKVSDLVANEFKDTQSTTV 440
DB 352 NTKSEELIGKNVSVFGYLLPEETIEI-----ITNIRADDIINTANKIFSQTTLAI 403
RESULT 3
YMXG_BACSU STANDARD; PRT; 409 AA.
AC 004805;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical zinc protease ymxg (EC 3.4.99.-) (ORFp).
GN YMXG.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=168 / 865;
RA Bolhuis A., Venema G., Bron S., Van Dijk J.M.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE OF 30-409 FROM N.A.
RC STRAIN=168;
RX MEDLINE=93252813; PubMed=8098035;
RA Chen N.-Y., Jiang S.-O., Klein D.A., Paulus H.;
RT "Organization and nucleotide sequence of the Bacillus subtilis
RT diaminopimelate operon, a cluster of genes encoding the first three
RT enzymes of diaminopimelate synthesis and dipicolinate synthase";
RL J. Biol. Chem. 268:9448-9465(1993).-----

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 11:53:44 ; Search time 14 Seconds
(without alignments)
1315.392 Million cell updates/sec

Title: US-09-881-752a-212

Perfect score: 2270
Sequence: 1 MKHFSVKRLGLSSVLVTL.....VRVANEYFKDTOSTTFLKP 444

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query length	ID	Description
1	511.5	22.5	512	Y4WA_RHISN	P55679 rhizobium s
2	340.5	13.0	412	Y219_RICPR	O05945 rickettsia
3	300	13.2	409	YMXG_BACSV	O04805 bacillus su
4	299	13.2	445	YR82_MYCLE	O32965 mycobacteri
5	297	13.1	480	UCRL_BOVIN	P31800 bos taurus
6	283.5	12.5	438	YR82_MYCTU	O33324 mycobacteri
7	274	12.1	931	PQOL_ECOCI	P31828 escherichia
8	265	11.7	480	UCRL_HUMAN	P31930 homo sapien
9	261	11.5	480	UCRL_MOUSE	O9C213 mus musculu
10	259.5	11.4	462	MPBP_YEAST	P10507 saccharomyc
11	256.5	11.3	459	YLD8_STRCO	O86835 streptomyc
12	253.5	11.2	471	YMTL_CAEEL	P98080 caenorhabdi
13	241.5	10.6	489	MPBP_MOUSE	O9C218 mus musculu
14	240	10.6	489	MPBP_RAT	O03346 rattus norv
15	239	10.5	926	POOL_HAEIN	P45181 haemophilus
16	234	10.3	489	MPBP_HUMAN	O75439 homo sapien
17	226.5	10.0	494	UCRL_EUGGR	P43264 euglena gra
18	220.5	9.7	476	MPBP_NEUCR	P11913 neurospora
19	213.5	9.4	524	MPBP_MOUSE	O9C261 mus musculu
20	213	9.4	482	MPBP_YEAST	P11914 saccharomyc
21	208	9.2	577	MPBP_NEUCR	P23955 neurospora
22	206	9.1	969	YEAC_SCHPO	O14077 schizosacch
23	205.5	9.1	525	MPBP_HUMAN	O10713 homo sapien
24	194.5	8.6	524	MPBP_RAT	P20069 rattus norv
25	194.5	8.6	973	HHPA_CLOPE	O46205 clostridium
26	185.5	8.2	1019	IDE_RAT	P35559 rattus norv
27	178.5	7.9	1036	YAN2_SCHPO	O10068 schizosacch
28	178	7.8	1018	IDE_HUMAN	P14735 homo sapien
29	175.5	7.7	1019	IDE_MOUSE	O9187 mus musculu
30	169	7.4	596	SDP_ETMBO	P42789 elmeria bov
31	160	7.0	988	STZ3_YEAST	O05010 saccharomyc
32	158.5	7.0	495	YHUT_SALTY	P50335 salmonella
33	155.5	6.9	498	YHUT_ECOCI	P37648 escherichia

34	152.5	6.7	457	1	UCRL_YEAST	P07256 saccharomyc
35	148.5	6.5	504	1	MPBP_SOLIU	P29677 solianum tub
36	147	6.5	426	1	UCR2_SCHPO	P78761 schizosacch
37	147	6.5	453	1	UCR2_HUMAN	P22695 homo sapien
38	146	6.4	447	1	Y4WB_RHISN	P55680 rhizobium s
39	146	6.4	447	1	PQGF_PSEAE	O912d2 pseudomonas
40	145.5	6.4	745	1	YQ4_CAEEL	O10040 caenorhabdi
41	142	6.3	1037	1	YQJ8_YEAST	O12496 saccharomyc
42	141.5	6.2	452	1	UCR2_YEAST	P32551 rattus norv
43	141.5	6.2	453	1	UCR2_MOUSE	O9db77 mus musculu
44	141	6.2	962	1	PTRA_ECOCI	P05458 escherichia
45	140.5	6.2	882	1	YBAH_SCHPO	O42908 schizosacch

ALIGNMENTS

RESULT 1	ID	Y4WA_RHISN	STANDARD:	PRT:	512 AA.
AC	P55679:				
DT	01-NOV-1997 (rel. 35, Last sequence update)				
DT	01-NOV-1997 (rel. 35, Last sequence update)				
DT	15-JUN-2002 (rel. 41, Last annotation update)				
DE	Hypothetical zinc protease y4wa (EC 3.4.99.-)				
GN	Y4WA.				
OS	Rhizobium sp. (strain NGR234).				
OC	Plasmid sym pNGR234.				
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;				
OC	Rhizobiaceae; Rhizobium.				
OX	NCBI_TaxID=394;				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE-97305956; PubMed-9163424;				
RA	Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,				
RA	Perret X.;				
RT	"Molecular basis of symbiosis between Rhizobium and legumes.";				
RL	Nature 387:394-401(1997).				
CC	-1 COFACTOR: REQUIRES DIVALENT CATIONS FOR ACTIVITY. BINDS ZINC (BY SIMILARITY).				
CC	-1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.				
CC	-1 SIMILARITY: TO Y4WB.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).				
CC	EMBL: AE00103; AAB91908.1; -.				
DR	MEROPS: M16.DPB; -.				
DR	InterPro: IPR001431; Peptidase_M16.				
DR	Pfam: PF00675; Peptidase_M16; 1.				
DR	PROSITE: PS00143; INSULINASE; 1.				
KW	Hypothetical protein; Hydrolase; Metalloprotease; zinc;				
KW	Transmembrane; Plasmid.				
FT	TRANSMEM 57 77				
FT	METAL 131 131				
FT	ACT_SITE 134 134				
FT	ACT_SITE 135 135				
FT	METAL 211 211				
FT	SEQUENCE 512 AA; 56886 MW; 7BDC60C11F08BD85 CRC64;				
QY	Query Match	22.5%;	Score 511.5;	DB 1;	Length 512;
DB	Best Local Similarity	32.1%;	Pred. No. 4.5e-25;		
DB	Matches 143;	Conservative 73;	Mismatches 212;	Indels 17;	Gaps 8;
DB	12 LSSVLTITGASMAHASTYLPKHE-----SVTLKNGQVVSPLENKTGYEDVLYKVGSR 67				
DB	62 LCMVALQFLMTSAMADESPLEAEVAVNMLGNGMEVVIP-DHRAPVTQMIWYKVGNA 120				

THIS PAGE BLANK (USPTO)

Qy 360 EKLKGEI 367
| | | | |
Db 468 EKLKGEI 475

Search completed: November 18, 2002, 12:04:55
job time : 26 secs

Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A:Title: Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti*
A:Reference number: A95262; MUID:21396509; PMID:11481432
A:Accession: F95363
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-370 <RUR>
A:Cross-references: GB:AE006469; PIDN:AAK65472.1; PID:g14523942; GSPDB:GN00165
A:Experimental source: strain 1021, megaplasmid pSymA
R:Galbert, F.; Fins, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hepault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: Sma1495
A:Genome: Plasmid
C:Superfamily: serine-pyruvate aminotransferase
C:Keywords: aminotransferase
Query Match 1.8%; Score 8; DB 2; Length 370;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 297 LSRLGEG 304
DB 262 LSRLGEG 269
RESULT 12
C75221
hypothetical protein PAB2152 - *Pyrococcus abyssi* (strain Orsay)
C:Species: *Pyrococcus abyssi*
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: C75221
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: C75221
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-385 <RAM>
A:Cross-references: GB:AJ248283; GB:AL096836; MID:95457433; PIDN:CAB49218.1; PID:e151511
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB2152
Query Match 1.8%; Score 8; DB 2; Length 385;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 295 DALSRLG 302
DB 340 DALSRLG 347
RESULT 13
T28278
ORF MSV117 probable DNA polymerase beta/AP endonuclease - *Melanoplus sanguinipes entomop*
C:Species: *Melanoplus sanguinipes entomopoxvirus*
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T28278
R:Alonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A:Title: The genome of *Melanoplus sanguinipes entomopoxvirus*.
A:Reference number: Z20484; MUID:99102612; PMID:9847359
A:Accession: T28278
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-603 <AFO>
A:Cross-references: EMBL:AF063866; MID:g4049647; PIDN:AAC97660.1; PID:g4049700
C:Genetics:
A:Note: MSV117
Query Match 1.8%; Score 8; DB 2; Length 603;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 383 FISLWESS 390
DB 445 FISLWESS 452
RESULT 14
G64336
large helicase related protein LHR homolog - *Methanococcus jannaschii*
C:Species: *Methanococcus jannaschii*
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 02-Feb-2001
C:Accession: G64336
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
Reich, C.T.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
Ison, J.D.; Sadov, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannasc*
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: G64336
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-841 <BUL>
A:Cross-references: GB:U67484; GB:L77117; MID:g1591009; PIDN:AAB98279.1; PID:g1591016
C:Genetics:
A:Map position: REV279066-276541
C:Keywords: ATP; nucleotide binding; P-loop
F:56-65/Region: nucleotide-binding motif A (P-loop)
F:177-182/Region: nucleotide-binding motif B
F:181-184/Region: DEXH motif
Query Match 1.8%; Score 8; DB 2; Length 841;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 360 EKLKKEI 367
DB 323 EKLKKEI 330
RESULT 15
F72253
hypothetical protein TM1450 - *Thermotoga maritima* (strain MSB8)
C:Species: *Thermotoga maritima*
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: F72253
R:Nelson, K.E.; Claydon, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: F72253
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-893 <ARN>
A:Cross-references: GB:AE001796; GB:AE000512; MID:g4982004; PIDN:AAD36518.1; PID:g498
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1450
Query Match 1.8%; Score 8; DB 2; Length 893;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```
OY 421 VSDLVRYANEYFKDOSTVFLKRP 444
DB 421 VSDLVRYANEYFKDOSTVFLKRP 444

RESULT 2
probable zinc proteinase - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 17-Mar-2000
C:Accession: D71935
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MID:99120557; PMID:9923682
A:Accession: D71935
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-443 <ARN>
A:Cross-references: GB:AE001475; GB:AE001439; NID:g4154939; PIDN:AAD05993.1; PID:g415494
A:Experimental source: strain J99
C:Genetics:
A:Gene: jhp0411
C:Superfamily: mitochondrial processing peptidase alpha chain

Query Match 14.4%; Score 64; DB 2; Length 443;
Best Local Similarity 100.0%; Pred. No. 3.6e-56;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 235 KKHESLKNDEKAIPYPYKKEPKODGARAVVHKDGVHLEWVALGYKPAFKHKDVAL 294
DB 234 KKHESLKNDEKAIPYPYKKEPKODGARAVVHKDGVHLEWVALGYKPAFKHKDVAL 293

OY 295 DALSL 298
DB 294 DALSL 297

RESULT 3
probable zinc proteinase Cj0805 [imported] - Campylobacter jejuni (strain NCTC 11168)
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: E81352
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyf
A:Reference number: A81250; MID:20150912; PMID:10688204
A:Accession: E81352
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-416 <PAR>
A:Cross-references: GB:AL139076; GB:AL111168; NID:g6968128; PIDN:CAB73070.1; PID:g696825
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cj0805
C:Superfamily: mitochondrial processing peptidase alpha chain

Query Match 6.1%; Score 27; DB 2; Length 416;
Best Local Similarity 100.0%; Pred. No. 8.8e-19;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 71 MGKSGIAHMLEHLNFKSTKNLKAEEFD 97
DB 42 MGKSGIAHMLEHLNFKSTKNLKAEEFD 68

RESULT 4
T16471
```

```
hypothetical protein F56C9.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T16471
R:Du, Z.
submitted to the EMBL Data Library, May 1994
A:Description: The sequence of C. elegans cosmid F56C9.
A:Reference number: S46729
A:Accession: T16471
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-237 <DUZ>
A:Cross-references: EMBL:U00063; NID:g488186; PID:g488193; PIDN:AAB08716.1; CESP:F56C
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:F56C9.7
A:Introns: 62/3; 103/2; 156/1

Query Match 2.0%; Score 9; DB 2; Length 237;
Best Local Similarity 100.0%; Pred. No. 0.85;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 GLSSVLLVLT 19
DB 8 GLSSVLLVLT 16

RESULT 5
T42365
uroporphyrin-III C-methyltransferase homolog - fission yeast (Schizosaccharomyces pom
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
C:Accession: T42365
R:Yoshioke, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.
DNA Res. 4, 363-369, 1997
A:Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.
A:Reference number: Z17323; MID:98162722; PMID:9501991
A:Accession: T42365
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-251 <YOS>
A:Cross-references: EMBL:DB9123; NID:g1749453; PIDN:BAI13785.1; PID:g1749454
A:Experimental source: strain PR745

Query Match 2.0%; Score 9; DB 2; Length 251;
Best Local Similarity 100.0%; Pred. No. 0.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 291 QVALDALSLR 299
DB 82 QVALDALSLR 90

RESULT 6
H64429
DNA-directed RNA polymerase (EC 2.7.7.6) subunit B' - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 08-Dec-2000
C:Accession: H64429
R:Butt, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
; Reich, C.J.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klek, H.P.; Fraser, C.M.; Smith, H.O.; Moese
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
A:Reference number: A64300; MID:96337999; PMID:8688087
A:Accession: H64429
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-636 <BUU>
A:Cross-references: GB:U67546; GB:L77117; NID:g1591687; PIDN:AAB99044.1; PID:g1591694
C:Genetics:
A:Map position: FOR972590-974500
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 12:02:25 ; Search time 22 Seconds
(without alignments)
1940.167 Million cell updates/sec

Title: US-09-881-752a-212

Perfect score: 444

Sequence: 1 MKHFSVKRLGLSSVLVLT.....VRVANEYFKDTQSTFVEIKP 444

Scoring table: OLIGO
Gapop 60.0 , Gapept 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	444	100.0	444	2 D64646	protease (EC 3.4
2	64	14.4	443	2 D71935	probable zinc prot
3	27	6.1	416	2 E81352	probable zinc prot
4	9	2.0	237	2 T16471	hypothetical prote
5	9	2.0	251	2 T42365	uroporphyrin-III C
6	9	2.0	636	2 H64429	DNA-directed RNA p
7	8	1.8	109	2 A17029	hypothetical prote
8	8	1.8	160	2 F81271	crossover junction
9	8	1.8	321	2 H96181	transaldolase taib
10	8	1.8	321	2 AD3105	transaldolase taib
11	8	1.8	370	2 F95363	probable serine-ly
12	8	1.8	385	2 C75221	hypothetical prote
13	8	1.8	603	2 T28278	hypothetical prote
14	8	1.8	841	2 G64336	large helicase rel
15	8	1.8	893	2 F72253	hypothetical prote
16	8	1.8	952	2 C82309	probable insulinas
17	7	1.6	113	2 T50077	hypothetical prote
18	7	1.6	117	2 A83312	hypothetical prote
19	7	1.6	139	2 S01156	conserved hypothet
20	7	1.6	141	2 S14259	Ig heavy chain pre
21	7	1.6	146	1 H8Y	seed storage prote
22	7	1.6	146	1 S55248	hemoglobin beta ch
23	7	1.6	147	1 H8Y	hemoglobin beta ch
24	7	1.6	148	2 H90424	hemoglobin beta ch
25	7	1.6	155	2 A95993	conserved hypothet
26	7	1.6	160	2 AC2067	vesa-like (mycopla
27	7	1.6	166	2 A81853	hypothetical prote
28	7	1.6	173	1 CFYCB	chaperone protein
29	7	1.6	189	2 C75252	C-phycocyanin beta
					hypothetical prote

30	7	1.6	194	1 Y0EC7P	fimbrial protein 9
31	7	1.6	198	2 AH0778	probable membrane
32	7	1.6	198	2 AC1364	protein gp51 [Bact
33	7	1.6	206	2 F71157	hypothetical prote
34	7	1.6	220	1 B42725	nitrite hydratase
35	7	1.6	223	2 B64396	hypothetical prote
36	7	1.6	224	2 H86117	hypothetical prote
37	7	1.6	224	2 S56431	hypothetical prote
38	7	1.6	224	2 H91276	hypothetical prote
39	7	1.6	229	2 F83663	serine O-acetyltra
40	7	1.6	232	2 D75062	probable flagella
41	7	1.6	232	2 H71169	hypothetical prote
42	7	1.6	232	2 F83911	transcription regu
43	7	1.6	234	2 E83024	NAD(P)H quinone ox
44	7	1.6	245	2 G64210	uracil DNA glycosy
45	7	1.6	249	2 H97251	hypothetical prote

ALIGNMENTS

RESULT 1	
D64646	
protease (EC 3.4.21.1) - Helicobacter pylori (strain 26695)	
C:Species: Helicobacter pylori	
C:Date: 09-Aug-1997 #sequence_revision 15-Aug-1997 #text_change 17-Mar-2000	
C:Accession: D64646	
R:Romd, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.	
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodek, A.; McK	
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,	
Nature 388, 539-547, 1997	
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,	
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.	
A:Reference number: A64520; MUID:97394467; PMID:9252185	
A:Accession: D64646	
A:Status: preliminary; nucleic acid sequence not shown; translation not shown	
A:Molecule type: DNA	
A:Residues: 1-444 <TOM>	
A:Cross-references: GB:AE00609; GB:AE00511; NID:g2314150; PIDN:AAD08056.1; PID:g231	
C:Superfamily: mitochondrial processing peptidase alpha chain	
C:Keywords: hydrolase	
Query Match	100.0%; Score 444; DB 2; Length 444;
Best local similarity	100.0%; Pred. No. 0;
Matches 444; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 1 MKHFSVKRLGLSSVLVLTGLGASMAQSYLPKHESVTLKNGLQVNVSPLENKTVIEVDV 60	
DB 1 MKHFSVKRLGLSSVLVLTGLGASMAQSYLPKHESVTLKNGLQVNVSPLENKTVIEVDV 60	
QY 61 LYKVGSRNFTMGKSGIAHLEHLNFKSTYKNTLAAGEFDKTVKFGVSNASTSPDITRYPI 120	
DB 61 LYKVGSRNFTMGKSGIAHLEHLNFKSTYKNTLAAGEFDKTVKFGVSNASTSPDITRYPI 120	
QY 121 KTSQANLDSLELFAETMGSLNLTKEDEFLPEROVVAEERWMTDNPISMLTFFREFNTAY 180	
DB 121 KTSQANLDSLELFAETMGSLNLTKEDEFLPEROVVAEERWMTDNPISMLTFFREFNTAY 180	
QY 181 VYHRYHWPDIQGMDDIQNTLTDIKKFHSLYYQPKNAVLVVGDVNSQKVFELSKKHFS 240	
DB 181 VYHRYHWPDIQGMDDIQNTLTDIKKFHSLYYQPKNAVLVVGDVNSQKVFELSKKHFS 240	
QY 241 LKNDLDEKALPTPYMKPEPKDGGARTAVVNDGVHLEVALGTVPAKHKDQVALDALSL 300	
DB 241 LKNDLDEKALPTPYMKPEPKDGGARTAVVNDGVHLEVALGTVPAKHKDQVALDALSL 300	
QY 301 LEEGSSNLSQSELYVKKRIASQAFSHNMOLQDESVFLTAGGNPNVKAALKEITVALL 360	
DB 301 LEEGSSNLSQSELYVKKRIASQAFSHNMOLQDESVFLTAGGNPNVKAALKEITVALL 360	
QY 361 KLRKEITQAEIDDKLKNKADFISNLESSSDVAGLFADYLVQNDIQGLTDVQROFLDK 420	
DB 361 KLRKEITQAEIDDKLKNKADFISNLESSSDVAGLFADYLVQNDIQGLTDVQROFLDK 420	

THIS PAGE BLANK (USPTO)

RT "Complete genomic sequence of the filamentous nitrogen-fixing
cytobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL: AP003587; BAB73490.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 109 AA; 12167 MW; 6DDED5C096A5C7D8 CRC64;

Query Match 1.8%; Score 8; DB 16; Length 109;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 VKRLGLS 13
DB 2 VKRLGLS 9

RESULT 13

OYTGK7 PRELIMINARY; PRT; 122 AA.
AC OYTGK7;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NADH dehydrogenase subunit 2 (Fragment).
GN ND2.
OS Echinococcus multilocularis.
OC Mitochondrion.
OC Eukaryota; Metazoa; Platyhelminthes; Cestoda; Eucestoda;
OC Cyclophyllidae; Taeniidae; Echinococcus.
OX NCBI_TaxID=6211;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21036605; PubMed=11163447;
RA Nakao M., Sako Y., Yokoyama N., Fukunaga M., Ito A.;
RT "Mitochondrial genetic code in cestodes.";
RL Mol. Biochem. Parasitol. 111:415-424(2000).
DR EMBL: AB031282; BAA83527.1;
KW Mitochondrion.
FT NON_TER 122
SQ SEQUENCE 122 AA; 14255 MW; 86693DB5E28E303B CRC64;

Query Match 1.8%; Score 8; DB 8; Length 122;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 GLSSVLLV 18
DB 72 GLSSVLLV 79

RESULT 14

OYTGJ8 PRELIMINARY; PRT; 123 AA.
AC OYTGJ8;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NADH dehydrogenase subunit 2 (Fragment).
GN ND2.
OS Taenia saginata.
OS Mitochondrion.
OC Eukaryota; Metazoa; Platyhelminthes; Cestoda; Eucestoda;
OC Cyclophyllidae; Taeniidae; Taenia.
OX NCBI_TaxID=6206;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21036605; PubMed=11163447;
RA Nakao M., Sako Y., Yokoyama N., Fukunaga M., Ito A.;
RT "Mitochondrial genetic code in cestodes.";
RL Mol. Biochem. Parasitol. 111:415-424(2000).
DR EMBL: AB031285; BAA83536.1;
KW Mitochondrion.
FT NON_TER 123

SQ SEQUENCE 123 AA; 14297 MW; BA21D72DB5FAFF96 CRC64;

Query Match 1.8%; Score 8; DB 8; Length 123;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 GLSSVLLV 18
DB 73 GLSSVLLV 80

RESULT 15

OYTE77 PRELIMINARY; PRT; 293 AA.
AC OYTE77;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE NADH dehydrogenase subunit 2.
GN ND2.
OS Echinococcus multilocularis.
OC Mitochondrion.
OC Eukaryota; Metazoa; Platyhelminthes; Cestoda; Eucestoda;
OC Cyclophyllidae; Taeniidae; Echinococcus.
OX NCBI_TaxID=6211;
RN [1]
RP SEQUENCE FROM N.A.
RA Fukunaga M.;
RT "Echinococcus multilocularis mitochondrial DNA sequence.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB018440; BAA84932.1;
KW Mitochondrion.
SQ SEQUENCE 293 AA; 34474 MW; AFC9D6649A4A8666 CRC64;

Query Match 1.8%; Score 8; DB 8; Length 293;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 GLSSVLLV 18
DB 72 GLSSVLLV 79

Search completed: November 18, 2002, 12:04:27
Job time : 39 secs

RT reveals multiple transfers between archaea and bacteria."
RL Mol. Biol. Evol. 18:362-375(2001).
DR EMBL: AJ401011; CAC21219.1; "-."
DR InterPro: IPR002587; Inos-1-P_synth.
DR Pfam: PF01658; Inos-1-P_synth. 1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 335 AA; 36890 MW; B62E3931072609D9 CRC64;

Query Match 2.0%; Score 9; DB 2; Length 335;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 LEKIKGEI 367
|||||
DB 7 LEKIKGEI 15

RESULT 9

Q9EVU0 PRELIMINARY; PRT; 335 AA.
AC O9EVU0:
DT 01-MAR-2001 (TremBrel. 16, Created)
DT 01-MAR-2001 (TremBrel. 16, last sequence update)
DT 01-DEC-2001 (TremBrel. 19, last annotation update)
DE Myo-Inositol 1P synthase (Fragment).
GN INO1.
OS Thermotoga sp. R07.
OC Bacteria; Thermotogae; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=126738;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R07;
RX MEDLINE-21143063; PubMed-11230537;
RA Nesbo C.L., L'Harrison S., Stetter K.O., Doolittle W.F.;
RT "Phylogenetic analyses of 'archaeal' genes in Thermotoga maritima
reveals multiple transfers between archaea and bacteria."
RL Mol. Biol. Evol. 18:362-375(2001).
DR EMBL: AJ401012; CAC21220.1; "-."
DR InterPro: IPR002587; Inos-1-P_synth.
DR Pfam: PF01658; Inos-1-P_synth. 1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 335 AA; 36618 MW; E6C6010B158EF511 CRC64;

Query Match 2.0%; Score 9; DB 2; Length 335;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 LEKIKGEI 367
|||||
DB 7 LEKIKGEI 15

RESULT 10

Q9EVU7 PRELIMINARY; PRT; 335 AA.
AC O9EVU7:
DT 01-MAR-2001 (TremBrel. 16, Created)
DT 01-MAR-2001 (TremBrel. 16, last sequence update)
DT 01-DEC-2001 (TremBrel. 19, last annotation update)
DE Myo-Inositol 1P synthase (Fragment).
GN INO1.
OS Thermotoga neapolitana.
OC Bacteria; Thermotogae; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2337;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LA10;
RX MEDLINE-21143063; PubMed-11230537;
RA Nesbo C.L., L'Harrison S., Stetter K.O., Doolittle W.F.;
RT "Phylogenetic analyses of 'archaeal' genes in Thermotoga maritima
reveals multiple transfers between archaea and bacteria."
RT

RL Mol. Biol. Evol. 18:362-375(2001).
DR EMBL: AJ401014; CAC21211.1; "-."
DR InterPro: IPR002587; Inos-1-P_synth.
DR Pfam: PF01658; Inos-1-P_synth. 1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 335 AA; 36719 MW; 4486D0D090E3525C CRC64;

Query Match 2.0%; Score 9; DB 2; Length 335;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 LEKIKGEI 367
|||||
DB 7 LEKIKGEI 15

RESULT 11

Q9EVU6 PRELIMINARY; PRT; 335 AA.
AC O9EVU6:
DT 01-MAR-2001 (TremBrel. 16, Created)
DT 01-MAR-2001 (TremBrel. 16, last sequence update)
DT 01-DEC-2001 (TremBrel. 19, last annotation update)
DE Myo-Inositol 1P synthase (Fragment).
GN INO1.
OS Thermotoga neapolitana.
OC Bacteria; Thermotogae; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2337;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NES;
RX MEDLINE-21143063; PubMed-11230537;
RA Nesbo C.L., L'Harrison S., Stetter K.O., Doolittle W.F.;
RT "Phylogenetic analyses of 'archaeal' genes in Thermotoga maritima
reveals multiple transfers between archaea and bacteria."
RL Mol. Biol. Evol. 18:362-375(2001).
DR EMBL: AJ401015; CAC21212.1; "-."
DR InterPro: IPR002587; Inos-1-P_synth.
DR Pfam: PF01658; Inos-1-P_synth. 1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 335 AA; 36394 MW; EE46DC9CFB07B13 CRC64;

Query Match 2.0%; Score 9; DB 2; Length 335;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 LEKIKGEI 367
|||||
DB 7 LEKIKGEI 15

RESULT 12

Q8YW25 PRELIMINARY; PRT; 109 AA.
AC Q8YW25:
DT 01-MAR-2002 (TremBrel. 20, Created)
DT 01-MAR-2002 (TremBrel. 20, last sequence update)
DT 01-MAR-2002 (TremBrel. 20, last annotation update)
DE Hypothetical protein Alr1791.
GN Alr1791.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE-21595285; PubMed-11759840;
RX Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida T., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;

RA Du 2.;
 RT "The sequence of C. elegans cosmid F56C9."
 RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RT "Direct Submission."
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U00063; AK18962.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 237 AA; 26300 MW; 3FED5E7668197A87 CRC64;

Query Match 2.0%; Score 9; DB 5; Length 237;
 Best Local Similarity 100.0%; Pred. No. 2.3;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GLSSVLLVT 19
 DB 8 GLSSVLLVT 16

RESULT 5

ID Q9EV08 PRELIMINARY; PRT; 335 AA.
 AC Q9EV08;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE Myo-Inositol 1P synthase (Fragment).
 GN INO1.
 OS Thermotoga neapolitana.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 OX NCBI_TaxID=2337;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP MEDLINE-21143063; PubMed-11230537;
 RA Nesbo C.L., L'Haridon S., Stetter K.O., Doolittle W.F.;
 RT "Phylogenetic analyses of 'archaeal' genes in Thermotoga maritima reveals multiple transfers between archaea and bacteria."
 RL MOL. Biol. Evol. 18:362-375(2001).
 DR EMBL: AJ401008; CAC21210.1; -
 DR InterPro: IPR002587; Inos-1-P_synth.
 DR Pfam: PF01658; Inos-1-P_synth; 1.
 FT NON_TER 1 1
 FT NON_TER 335 335
 SQ SEQUENCE 335 AA; 36560 MW; C44AD10184ABF110 CRC64;

Query Match 2.0%; Score 9; DB 2; Length 335;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 LEKLKGEI 367
 DB 7 LEKLKGEI 15

RESULT 6

ID Q9EV01 PRELIMINARY; PRT; 335 AA.
 AC Q9EV01;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE Myo-Inositol 1P synthase (Fragment).
 GN INO1.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP MEDLINE-21143063; PubMed-11230537;
 RA Nesbo C.L., L'Haridon S., Stetter K.O., Doolittle W.F.;

RT "Phylogenetic analyses of 'archaeal' genes in Thermotoga maritima reveals multiple transfers between archaea and bacteria."
 RL MOL. Biol. Evol. 18:362-375(2001).
 DR EMBL: AJ401009; CAC21206.1; -
 DR InterPro: IPR002587; Inos-1-P_synth.
 DR Pfam: PF01658; Inos-1-P_synth; 1.
 FT NON_TER 1 1
 FT NON_TER 335 335
 SQ SEQUENCE 335 AA; 36611 MW; 1DEECAS80FAF6A84 CRC64;

Query Match 2.0%; Score 9; DB 2; Length 335;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 LEKLKGEI 367
 DB 7 LEKLKGEI 15

RESULT 7

ID Q9EV00 PRELIMINARY; PRT; 335 AA.
 AC Q9EV00;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE Myo-Inositol 1P synthase (Fragment).
 GN INO1.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN-SL7;
 RP MEDLINE-21143063; PubMed-11230537;
 RA Nesbo C.L., L'Haridon S., Stetter K.O., Doolittle W.F.;
 RT "Phylogenetic analyses of 'archaeal' genes in Thermotoga maritima reveals multiple transfers between archaea and bacteria."
 RL MOL. Biol. Evol. 18:362-375(2001).
 DR EMBL: AJ401010; CAC21207.1; -
 DR InterPro: IPR002587; Inos-1-P_synth.
 DR Pfam: PF01658; Inos-1-P_synth; 1.
 FT NON_TER 1 1
 FT NON_TER 335 335
 SQ SEQUENCE 335 AA; 36646 MW; AF9EALC5D5C48014 CRC64;

Query Match 2.0%; Score 9; DB 2; Length 335;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 LEKLKGEI 367
 DB 7 LEKLKGEI 15

RESULT 8

ID Q9EV01 PRELIMINARY; PRT; 335 AA.
 AC Q9EV01;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE Myo-Inositol 1P synthase (Fragment).
 GN INO1.
 OS Thermotoga sp. KO16.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 OX NCBI_TaxID=126741;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN-KOL6;
 RP MEDLINE-21143063; PubMed-11230537;
 RA Nesbo C.L., L'Haridon S., Stetter K.O., Doolittle W.F.;
 RT "Phylogenetic analyses of 'archaeal' genes in Thermotoga maritima

Query Match 100.0%; Score 444; DB 16; Length 444;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 444; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKHFSVRLGLSSVLTGASMHAAOSYLPKHESVTLKKNLOVYVPLEKKTGVIEVDY 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1 MKHFSVRLGLSSVLTGASMHAAOSYLPKHESVTLKKNLOVYVPLEKKTGVIEVDY 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 61 LYKGSRNRTGKSGIAHMLHFNKSTKNLAGEFDKIVRFGGVASASTSPDITRFYI 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 61 LYKGSRNRTGKSGIAHMLHFNKSTKNLAGEFDKIVRFGGVASASTSPDITRFYI 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 121 KTSQANLDRSLLEFAETGWSLNKLEDEFLPERQVAAEERMRRTNSPIGMLYFRFNTAY 180
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 121 KTSQANLDRSLLEFAETGWSLNKLEDEFLPERQVAAEERMRRTNSPIGMLYFRFNTAY 180
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 181 VYHRYHWPPIGFMDIOWITLKDKKPHSLTYQPKNAIVLVGVNSQKVELSKKHES 240
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 181 VYHRYHWPPIGFMDIOWITLKDKKPHSLTYQPKNAIVLVGVNSQKVELSKKHES 240
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 241 LKNDEKAIPTPYMKPEKODGARTAVVHKDGVHLEWVALGYKVPFKHKKDOVALDALSR 300
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 241 LKNDEKAIPTPYMKPEKODGARTAVVHKDGVHLEWVALGYKVPFKHKKDOVALDALSR 300
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 301 LGESKSSWLOSELVDKRRLASQAFSHNQLODESVEFLFIAGNPNVKAALQKEIVALLE 360
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 301 LGESKSSWLOSELVDKRRLASQAFSHNQLODESVEFLFIAGNPNVKAALQKEIVALLE 360
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 361 KLKGGETQALDKLKINQKADFTSNLESSDVAGLPADYIVQNDIGCLDTYQROFLDLK 420
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 361 KLKGGETQALDKLKINQKADFTSNLESSDVAGLPADYIVQNDIGCLDTYQROFLDLK 420
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 421 VSDLVRAVNEFYKDTOSTTVEFLKP 444
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 421 VSDLVRAVNEFYKDTOSTTVEFLKP 444
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 2
 Q92M12 PRELIMINARY; PRT; 443 AA.
 ID 092M12; PRT; 443 AA.
 AC 092M12; PRT; 443 AA.
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Putative zinc protease.
 GN JHP0411.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 OX NCBI_TaxID=85963;
 RN 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
 RA Tummino P.J., Carnuso A., Uria-Nickelsen M., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
 RA Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen Helicobacter pylori.";
 RL Nature 387:176-180(1999).
 DR EMBL: AE001475; AAD05993.1; -
 DR InterPro: IPR001431; Peptidase_M16.
 DR Pfam: PF00675; Peptidase_M16; 1.
 DR PROSITE: PS00143; INSULINASE; 1.
 KW Complete proteome.
 SQ SEQUENCE 443 AA; 50032 MW; 861D38FCTE53FFB2 CRC64;

Query Match 14.4%; Score 64; DB 16; Length 443;
 Best Local Similarity 100.0%; Pred. No. 2e-55;
 Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 KKHESLKNLDEKAIPTPYMKPEKODGARTAVVHKDGVHLEWVALGYKVPFKHKKDOVAL 294

DB 234 KKHESLKNLDEKAIPTPYMKPEKODGARTAVVHKDGVHLEWVALGYKVPFKHKKDOVAL 293
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 295 DALS 298
 |||||
 DB 294 DALS 297

RESULT 3
 Q9PPB5 PRELIMINARY; PRT; 416 AA.
 ID 09PPB5; PRT; 416 AA.
 AC 09PPB5; PRT; 416 AA.
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Putative zinc protease.
 GN CU0805.
 OS Campylobacter jejuni.
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
 OC Campylobacter.
 OX NCBI_TaxID=197;
 RN 11
 RP SEQUENCE FROM N.A.
 RX STRAIN=NCTC 11168;
 RC MEDLINE=20150912; PubMed=10688204;
 RA Parkhill J., Wren B.W., Mungall K., Kelsey J.M., Churcher C.,
 RA Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,
 RA Jorgels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
 RA Whitehead S., Barrett B.G.;
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
 RT reveals hypervariable sequences.";
 RL Nature 403:665-668(2000).
 DR EMBL: AL139076; CAB73070.1; -
 DR InterPro: IPR001431; Peptidase_M16.
 DR Pfam: PF00675; Peptidase_M16; 1.
 DR PROSITE: PS00143; INSULINASE; 1.
 KW Complete proteome.
 SQ SEQUENCE 416 AA; 48043 MW; 9A40AC42050B0DCE CRC64;

Query Match 6.1%; Score 27; DB 16; Length 416;
 Best Local Similarity 100.0%; Pred. No. 3e-18;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 MGRSGIAHMLHFNKSTKNLAGEFD 97
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 42 MGRSGIAHMLHFNKSTKNLAGEFD 68
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 4
 Q20865 PRELIMINARY; PRT; 237 AA.
 ID 020865; PRT; 237 AA.
 AC 020865; PRT; 237 AA.
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Hypothetical 26.3 kDa protein.
 GN F56C9.7.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN 11
 RP SEQUENCE FROM N.A.
 RX STRAIN=BRISTOL N2;
 RC MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN 12
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 11:59:50 ; Search time 35 Seconds
(without alignments)
2613.855 Million cell updates/sec

Title: US-09-881-752a-212

Perfect score: 444
Sequence: 1 MKHFSVKRLGLSLVLTWLT.....VRANEFKDTOSTVFLKP 444

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp Unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	444	100.0	444	16	025656 helicobacte
2	64	14.4	443	16	092m12 helicobacte
3	27	6.1	416	16	09ppb5 campylobact
4	9	2.0	237	5	020865 caenorhabd
5	9	2.0	335	2	09EVU8 thermotoga
6	9	2.0	335	2	09EVU1 thermotoga
7	9	2.0	335	2	09EVU0 thermotoga
8	9	2.0	335	2	09EVU1 thermotoga
9	9	2.0	335	2	09EVU7 thermotoga
10	9	2.0	335	2	09EVU7 thermotoga
11	9	2.0	335	2	09EVU6 thermotoga
12	8	1.8	109	16	08Yw25 anabaena sp
13	8	1.8	122	8	09Tgk7 echinococcu
14	8	1.8	123	8	09TgJ8 taenia sagi
15	8	1.8	293	8	09TgE7 echinococcu
16	8	1.8	293	8	095814 hymenolepis

17	8	1.8	293	8	0953M6	0953M6 echinococcu
18	8	1.8	321	16	08U715	08U715 agrobacteri
19	8	1.8	345	16	08RHPS	08RHPS fusobacteri
20	8	1.8	370	16	092Yp9	092Yp9 rhizobium m
21	8	1.8	385	17	09Y1Y0	09Y1Y0 pyrococcus
22	8	1.8	462	4	09NVC3	09NVC3 homo sapien
23	8	1.8	552	10	094LG7	094LG7 oryza sativ
24	8	1.8	603	12	09YVX5	09YVX5 melanoplus
25	8	1.8	745	5	08WST7	08WST7 halocynthia
26	8	1.8	893	16	09XIG2	09XIG2 thermotoga
27	8	1.8	952	16	09KUG7	09KUG7 vibrio chol
28	8	1.8	1256	5	09W238	09W238 drosophila
29	8	1.8	2209	5	097324	097324 plasmodium
30	7	1.6	76	1	09HH96	09HH96 sulfolobus
31	7	1.6	90	2	09AJ96	09AJ96 streptococc
32	7	1.6	91	12	09JFX1	09JFX1 chilo iride
33	7	1.6	113	3	09URK1	09URK1 schizosacch
34	7	1.6	117	16	0910H1	0910H1 pseudomonas
35	7	1.6	122	8	09TgK4	09TgK4 echinococcu
36	7	1.6	122	8	09TgI6	09TgI6 mesocostoid
37	7	1.6	123	8	09TgK1	09TgK1 taenia soli
38	7	1.6	128	17	08U467	08U467 sulfolobus
39	7	1.6	148	17	097V75	097V75 mycoplasma
40	7	1.6	155	16	098P50	098P50 anabaena sp
41	7	1.6	160	16	08YV92	08YV92 neisseria m
42	7	1.6	166	16	09JYX6	09JYX6 neisseria m
43	7	1.6	166	16	09JTW6	09JTW6 oryza offic
44	7	1.6	173	10	09M5N1	09M5N1 oryza meyer
45	7	1.6	173	10	09FUR1	09FUR1 oryza meyer

ALIGNMENTS

RESULT 1
ID 025656 PRELIMINARY: PRT; 444 AA.
AC 025656;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Protease (P009).
GN HP1012.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26595 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Goeysne J.D., Uitterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Mallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori";
RL Nature 388:539-547(1997).
DR EMBL: AE000609; AAD08056.1; -.
DR TIGR: HP1012; -.
DR InterPro: IPR001431; Peptidase_M16.
DR InterPro: IPR003880; Papain_attach.
DR Pfam: PF00675; Peptidase_M16; 1.
DR PROSITE: PS00143; INSULINASE; 1.
DR PROSITE: PS00012; PHOSPHOTANTHEINE; UNKNOWN_1.
KW Hypothetical protein; Protease; Complete proteome.
SQ SEQUENCE 444 AA; 50330 MW; 3A6092EF0792EAB2 CRC64;

```

CC -----
DR EMBL; X96471; CAA65324.2; .
DR EMBL; AP005277; BAB98655.1; ALT_INIT.
DR InterPro: IPR001123; Lyse.
DR InterPro: IPR004777; Lys_exporter.
DR Pfam; PF01810; Lyse; 1.
DR TIGRFAMs; TIGR00948; Za75; 1.
KW Transport; Transmembrane; Inner membrane.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 35 55 POTENTIAL.
FT TRANSMEM 66 86 POTENTIAL.
FT TRANSMEM 144 164 POTENTIAL.
FT TRANSMEM 177 197 POTENTIAL.
FT TRANSMEM 213 233 POTENTIAL.
SQ SEQUENCE 233 AA; 25082 MW; F5FD9B1ACAD11D13 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 233;
- Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 335 VFLEFIAG 341
   |||||
FB 47 VFLEFIAG 53

```

Search completed: November 18, 2002, 12:03:44
 Job time : 17 secs

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AP000002; BAA29547.1; ALT_INIT.
 DR InterPro: IPR002845; DUF127.
 DR Pfam: PF01994; DUF127.1.
 DR ProDom: PD016584; DUF127.1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 203 AA; 22767 MW; E81C17F5848D3414 CRC64;
 Query Match 1.6%; Score 7; DB 1; Length 203;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 360 EKLKGE 366
 DB 96 EKLKGE 102
 RESULT 11
 YTFB_ECOLI
 ID YTFB_ECOLI STANDARD; PRT; 212 AA.
 AC P39310;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein ytfB.
 GN YTFB OR B4206.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 OC NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=95334362; PubMed=7610040;
 RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
 RA Blattner F.R.;
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
 RT region from 92.8 through 100 minutes."
 RL Nucleic Acids Res. 23:2105-2119(1995).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U14003; AAA97102.1; ALT_INIT.
 DR EMBL: AE000492; AAC72763.1; ALT_INIT.
 DR Ecogene: EG12502; ytfB.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 34 50 POTENTIAL.
 SQ SEQUENCE 212 AA; 23505 MW; 7E475BDE4933A164 CRC64;
 Query Match 1.6%; Score 7; DB 1; Length 212;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 217 AIVLVVG 223
 DB 39 AIVLVVG 45
 RESULT 12
 NHAB_PSECL
 ID NHAB_PSECL STANDARD; PRT; 220 AA.
 AC P27763;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Nitrile hydratase subunit beta (EC 4.2.1.84) (Nitrilase) (NHase).
 GN NHAB.
 OS Pseudomonas chlororaphis (Pseudomonas aureofaciens).
 CC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 CC Pseudomonas.
 OC NCBI_TaxID=333;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-23 AND 158-170.
 RC STRAIN-B23;
 RX MEDLINE=91193202; PubMed=2013568;
 RA Nishiyama M., Horinouchi S., Kobayashi M., Nagasawa T., Yamada H.,
 RA Beppu T.;
 RT "Cloning and characterization of genes responsible for metabolism of
 RT nitrile compounds from Pseudomonas chlororaphis B23.";
 RL J. Bacteriol. 173:2465-2472(1991).
 CC -!- FUNCTION: NHASE CATALYZES THE HYDRATION OF VARIOUS NITRILE
 CC COMPOUNDS TO THE CORRESPONDING AMIDES. INDUSTRIAL PRODUCTION OF
 CC ACRYLAMIDE IS NOW BEING DEVELOPED USING SOME OF THESE ENZYMES.
 CC -!- CATALYTIC ACTIVITY: An aliphatic amide = a nitrile + H(2)O.
 CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
 CC -!- SIMILARITY: BELONGS TO THE NITRILE HYDRATASE SUBUNIT BETA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D90216; BAA14246.1; -
 DR PIR: B42725; B42725.
 DR HSSP: P13449; 2AHJ.
 DR InterPro: IPR003168; NHase_Beta.
 DR Pfam: PF02211; NHase_Beta.1.
 KW Lyase.
 SQ SEQUENCE 220 AA; 24545 MW; 191AE1C5F14D4864 CRC64;
 Query Match 1.6%; Score 7; DB 1; Length 220;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 367 ITQAEID 373
 DB 91 ITQAEID 97
 RESULT 13
 Y770_METJA
 ID Y770_METJA STANDARD; PRT; 223 AA.
 AC O58180;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MJ0770.
 GN MJ0770.
 OS Methanococcus jannaschii.
 CC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 CC Methanocaldococcaceae; Methanocaldococcus.
 OC NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.U., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kierlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weissstock K.G., Merrick J.L., Nguyen D.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhman J.M., Nguyen D.,
 RA Ufferback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Bordovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;

RX MEDLINE-87174767; PubMed-3104880;
 RA Lau P.C.K.; Condie A.; Alvarado-Urbina G.; Lau R.H.;
 RT "Nucleotide sequence of phycocyanin beta-subunit gene of
 cyanobacterium *Anacystis nidulans* strain R2.";
 RL Nucleic Acids Res. 15:2394-2394(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-88257006; PubMed-2454910;
 RA Kalla R.S.; Lind L.K.; Lidholm J.; Gustafsson P.;
 RT "Transcriptional organization of the phycocyanin subunit gene
 clusters of the cyanobacterium *Anacystis nidulans* UTEX 625.";
 RL J. Bacteriol. 170:2961-2970(1988).
 RN [3]
 RP SEQUENCE OF 170-172 FROM N.A.
 RX MEDLINE-87248092; PubMed-3036657;
 RA Lau R.H.; Alvarado-Urbina G.; Lau P.C.K.;
 RT "Phycocyanin alpha-subunit gene of *Anacystis nidulans* R2: cloning,
 nucleotide sequencing and expression in *Escherichia coli*.";
 RL Gene 52:21-29(1987).
 CC -1- FUNCTION: LIGHT-HARVESTING PHOTOSYNTHETIC BILE PIGMENT-PROTEIN
 CC FROM THE PHYCOBILIPROTEIN COMPLEX.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
 CC -1- PTM: CONTAINS TWO COVALENTLY LINKED BILIN CHROMOPHORES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
 CC or send an email to license@slb.ch).
 CC -----
 DR EMBL: X04916; CAA28585.1; -
 DR EMBL: M94218; AAA64531.1; -
 DR EMBL: M16325; -; NOT_ANNOTATED_CDS.
 DR EMBL: M22706; AAA21760.1; -
 DR EMBL: M22704; AAA21759.1; -
 DR PIR: A26577; A26577.
 DR HSSP: P07119; 1CPC.
 DR InterPro: IPR001659; Phycobillosome.
 DR Pfam: PF00502; Phycobillosome; 1.
 DR ProDom: PD000340; Phycobillosome; 1.
 KW Phycobillosome; Electron transport; Photosynthesis; Bile pigment;
 KM Methylation.
 FT INIT MET 0 0 BY SIMILARITY.
 FT MOD.RES 72 72 METHYLATION (BY SIMILARITY).
 FT BINDING 82 82 PHYCOCYANOBILIN CHROMOPHORE.
 FT BINDING 153 153 PHYCOCYANOBILIN CHROMOPHORE.
 GN SEQUENCE 172 AA; 18138 MW; 4AC821C86D0BEA4D CRC64;
 SQ
 Query Match 1.6%; Score 7; DB 1; Length 172;
 * Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 294 LDALSRL 300
 DB 24 LDALSRL 30
 RESULT 9
 FM98_ECOLI STANDARD; PRT; 194 AA.
 ID FM98_ECOLI
 AC P21413;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Fimbrial protein 987p precursor (Fimbrial adhesin 987p).
 GN FASA OR FAPC.
 OS *Escherichia coli*.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC *Escherichia*.
 CC NCBI_TaxID=562;
 RN [1]

RP SEQUENCE FROM N.A.
 RA de Graaf F.K.; Klaassen P.;
 RT "The nucleotide sequence of the gene encoding the K99 subunit of
 enterotoxigenic *Escherichia coli*.";
 RL FEMS Microbiol. Lett. 42:253-258(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-987;
 RA Schifferli D.M.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE FIMB/PAPA FAMILY OF FIMBRIA PROTEINS.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-3 IS THE INITIATOR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
 CC or send an email to license@slb.ch).
 CC -----
 DR EMBL: M35257; AAA23405.1; -
 DR EMBL: U50547; AAB02684.1; -
 DR PIR: S06261; Y0EC7P.
 DR InterPro: IPR000259; Fimbrial.
 DR Pfam: PF00419; Fimbrial; 1.
 KW Fimbria; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 194 FIMBRIAL PROTEIN 987P.
 FT DISULFID 46 85 BY SIMILARITY.
 FT SEQUENCE 194 AA; 19660 MW; B2B5F7089584AD96 CRC64;
 SQ

Query Match 1.6%; Score 7; DB 1; Length 194;
 * Best Local Similarity 100.0%; Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 122 TSOAND 128
 DB 30 TSOAND 36
 RESULT 10
 ID Y461_PYRHO STANDARD; PRT; 203 AA.
 AC 058214;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein PH0461.
 GN PH0461.
 OS *Pyrococcus horikoshii*.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC *Pyrococcus*.
 OX NCBI_TaxID=53953;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OT3;
 RX MEDLINE-98344137; PubMed-9679194;
 RA Kawarabayashi Y.; Sawada M.; Horikawa H.; Hino Y.;
 RA Yamamoto S.; Sekine M.; Baba S.-I.; Kosugi H.; Hosoyama A.; Nagai Y.;
 RA Sakai M.; Ogura K.; Otsuka R.; Nakazawa H.; Takamiya M.; Ohfuku Y.;
 RA Funahashi T.; Tanaka T.; Kudoh Y.; Yamazaki J.; Kushida N.; Oguchi A.;
 RA Aoki K.-I.; Yoshizawa T.; Nakamura Y.; Robb F.T.; Horikoshi K.;
 RA Masuchi Y.; Shizuya H.; Kikuchi H.;
 RT "Complete sequence and gene organization of the genome of a hyper-
 RT thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3.";
 RL DNA Res. 5:55-76(1998).
 CC -1- SIMILARITY: BELONGS TO THE UPF0106 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X56686; CAA40015.1; -.
DR PIR: S14259; S14259.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; try_amy1_inhbt.
DR Pfam: PF00234; try_alpha_amy1.1.
DR SMART: SM00499; AAI; 1.
KW Seed storage protein; Albumin; signal; Multigene family.
FT SIGNAL 1 25
FT PROPEP 26 38
FT CHAIN 39 141 ALBUMIN 8.
FT DISULFID 49 100 POTENTIAL.
FT DISULFID 62 89 POTENTIAL.
FT DISULFID 90 132 POTENTIAL.
FT DISULFID 102 139 POTENTIAL.
FT CONFLICT 67 67 M -> N (IN REF. 2).
SQ SEQUENCE 141 AA; 16090 MW; 1E5723B9122C9BD4 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 141;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 LKNDK 247
Db 93 LKNDK 99

RESULT 6
HBB2_XENTR
ID HBB2_XENTR STANDARD; PRT; 146 AA.
AC P08423;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hemoglobin beta-2 chain (larval).
OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodidae; Silurana.
OX NCBI_TaxID=8364;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-88096508; Pubmed-3697074;
RA Knoechel W., Beck J., Meyerhof W.;
RT "Nucleotide sequence of the Xenopus tropicalis larval beta globin
RT gene.";
RL Nucleic Acids Res. 15:10062-10062(1987).
CC -1- FUNCTION: THIS IS A LARVAL (TADPOLE) BETA-GLOBIN.
CC -1- SUBUNIT: HETEROTETRAMER OF TWO ALPHA CHAINS AND TWO BETA CHAINS.
CC -1- TISSUE SPECIFICITY: RED BLOOD CELLS.
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y00501; CAA68554.1; -.
DR PIR: S06308; HBXLW.
DR HSSP: P02096; IFDH.
DR InterPro: IPR000971; Globin.
DR Pfam: PF00042; globin.1.
DR PROSITE: PS01033; GLOBIN; 1.

```

```

KW Heme; Oxygen transport; Transport; Erythrocyte.
FT INT MET 0 0
FT METAL 63 63 IRON (HEME DISTAL LIGAND).
FT METAL 92 92 IRON (HEME PROXIMAL LIGAND).
SQ SEQUENCE 146 AA; 15869 MW; 61CC8AE0AFC160E CRC64;

Query Match 1.6%; Score 7; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 295 DALSRLL 301
Db 26 DALSRLL 32

RESULT 7
HBB_COLLI
ID HBB_COLLI STANDARD; PRT; 146 AA.
AC P11342;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemoglobin beta chain.
KW HBB.
OS Columba livia (Domestic pigeon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Columbiformes; Columbidae; Columba.
OX NCBI_TaxID=8932;
RN [1]
RP SEQUENCE.
RA Zaidi Z.H.;
RL Submitted (MAY-1988) to the PIR data bank.
CC -1- FUNCTION: INVOLVED IN OXYGEN TRANSPORT FROM THE LUNG TO THE
CC VARIOUS PERIPHERAL TISSUES.
CC -1- SUBUNIT: HETEROTETRAMER OF TWO ALPHA CHAINS AND TWO BETA CHAINS.
CC -1- TISSUE SPECIFICITY: RED BLOOD CELLS.
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
DR PIR: J00340; J00340.
DR PIR: J00340; J00340.
DR HSSP: P02118; 1A4F.
DR InterPro: IPR002337; Beta_haem.
DR InterPro: IPR000971; Globin.
DR Pfam: PF00042; globin.1.
DR PRINTS: PR00814; BETAHAEM.
DR PROSITE: PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Erythrocyte.
FT METAL 63 63 IRON (HEME DISTAL LIGAND).
FT METAL 92 92 IRON (HEME PROXIMAL LIGAND).
SQ SEQUENCE 146 AA; 16152 MW; 3F9698269D2F06FD CRC64;

Query Match 1.6%; Score 7; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 342 GNPVKA 348
Db 56 GNPVKA 62

RESULT 8
PHCB_SYNP7
ID PHCB_SYNP7 STANDARD; PRT; 172 AA.
AC P06539; 057358;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE C-phycocyanin beta chain.
GN PCRB OR CPCL.
OS Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=1140;
RN [1]
RP SEQUENCE FROM N.A.

```

AC 014833:;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Metabotropic glutamate receptor 4 precursor (mglur4).
GN GRM4 OR GPRC4 OR MGLUR4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96346635, PubMed=8738157;
RA Makoff A., Leclerc R., Oxer M., Harrington K., Emson P.;
RT "Molecular characterization and localization of human metabotropic
RL glutamate receptor type 4.";
RL Brain Res. Mol. Brain Res. 37:239-248(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98141892; PubMed=9473604;
RA Wu S., Wright R.A., Rockey P.K., Burgett S.G., Arnold J.S.,
RT Rostek P.R. Jr., Johnson B.G., Schepp D.D., Belagaje R.M.;
RT "Group III human metabotropic glutamate receptors 4, 7 and 8:
RT molecular cloning, functional expression, and comparison of
RL pharmacological properties in RGT cells.";
RL Brain Res. Mol. Brain Res. 53:88-97(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95342351; PubMed=7617140;
RA Flor P.J., Lukic S., Rueegg D., Leonhardt T., Knoepfel T., Kuhn R.;
RT "Molecular cloning, functional expression and pharmacological
RT characterization of the human metabotropic glutamate receptor type
RT 4.";
RL Neuropharmacology 34:149-155(1995).
RN [4]
RP VARIANT ILE-797.
RX MEDLINE=21416233; PubMed=11525421;
RA Ohtsuki T., Toru M., Arinami T.;
RT "Mutation screening of the metabotropic glutamate receptor mglur4
RT (GRM4) gene in patients with schizophrenia.";
RL Psychiatr. Genet. 11:79-83(2001).
CC -1- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR
CC IS MEDIATED BY A G-PROTEIN THAT INHIBITS ADENYLATE CYCLASE
CC ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN THE CEREBELLUM.
CC EXPRESSED AT LOW LEVELS IN HIPPOCAMPUS, HYPOTHALAMUS AND
CC THALAMUS. NO EXPRESSION DETECTED IN LIVER.
CC -1- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC STRONGEST, TO MGLUR6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X80818; CAAS6784.1; -;
DR EMBL; U92457; AAB51762.1; -;
DR Genew; HGNC:4596; GRM4.
DR MIM; 604100; -;
DR InterPro; IPR001828; ANF_receptor.
DR InterPro; IPR000337; GPCR_Mgr.
DR Pfam; PF00003; 7tm_3; 1.
DR Pfam; PF01094; ANF_receptor; 1.
DR PRINTS; PR00248; GPCRMR.
DR PROSITE; PS00979; G_PROTEIN_REC_P3_1; 1.
DR PROSITE; PS00980; G_PROTEIN_REC_P3_2; 1.
DR PROSITE; PS00981; G_PROTEIN_REC_P3_3; 1.

KW	G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KM	Multigene family; Polymorphism.
FT	SIGNAL
FT	CHAIN
FT	DOMAIN
FT	TRANSSEM
FT	DOMAIN
FT	TRANSSEM
FT	DOMAIN
FT	TRANSSEM
FT	DOMAIN
FT	TRANSSEM
FT	DOMAIN
FT	TRANSSEM
FT	DOMAIN
FT	CARBOHYD
FT	CARBOHYD
FT	CARBOHYD
FT	CARBOHYD
FT	VARIANT
SO	SEQUENCE
QY	92 KAGEFDKI 99
Db	261 KAGEFDKI 268
RESULT 5	
ID	2SS8_HELAN
AC	P23110:
DT	01-NOV-1991 (Rel. 20, Created)
DT	01-NOV-1991 (Rel. 20, Last sequence update)
DE	Albumin 8 precursor (methionine-rich 2S protein) (SFAS).
OS	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta; Spermatophyta: Magnoliophyta; eudicotyledons; core eudicots; Asteridae: euasterids II; Asterales; Asteraceae; Asteroideae; Heliantheae; Helianthus.
OC	Asteridae: euasterids II; Asterales; Asteraceae; Asteroideae; Heliantheae; Helianthus.
OX	NCBI_TaxID=4232;
RN	[1]
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 39-141.
RC	TISSUE=Seed;
RX	MEDLINE=91146568; PubMed=1997318;
RA	Koritz A.A., Caldwell J.B., Lilley G.G., Higgins T.J.V.;
RT	"Amino acid and cDNA sequences of a methionine-rich 2S protein from sunflower seed (Helianthus annuus L.).";
RL	Eur. J. Biochem. 195:329-334(1991).
RN	[2]
RP	SEQUENCE OF 39-141.
RC	STRAIN=cv. Hybrid 246; TISSUE=Seed;
RX	MEDLINE=97072195; PubMed=8915004;
RA	Egorov T.A., Odintsova T.I., Musolymov A.K., Fido R., Tatham A.S., Shewry P.R.;
RT	"Disulphide structure of a sunflower seed albumin: conserved and variant disulphide bonds in the cereal prolamin superfamily.";
RL	FEBS Lett. 396:285-288(1996).
CC	-1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC	-1- SUBUNIT: DISULFIDE LINKED HEMEROIMER.
CC	-1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC	-----

SQ SEQUENCE 636 AA; 72193 MW; 81DE30267A572914 CRC64;

Query Match 2.0%; Score 9; DB 1; Length 636;
Best Local Similarity 100.0%; Pred. No. 0.77;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 360 EKLKGEIT 368
IIIIIIII
DB 84 EKLKGEIT 92

RESULT 2

RUV_CAMJE STANDARD; PRT; 160 AA.

ID RUV_CAMJE

AC 09PLU8;

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DE Crossover junction endonuclease ruvc (EC 3.1.22.4) (Holliday

junction nuclelease ruvc) (Holliday junction resolvase ruvc).

GN RUV_C OR C11731C.

OS Campylobacter jejuni.

OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;

OC Campylobacter.

OX NCBI_TaxID=197;

RA [1]

RP SEQUENCE FROM N.A.

RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;

RX MEDLINE=96337999; PubMed=8688087;

RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,

RA Overbeek R., Kirkness E.F., Weissstock K.G., Merrick J.M., Glodok A.,

RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,

RA Utechak T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,

RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,

RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;

RT "Complete genome sequence of the methanogenic archaeon, Methanococcus

RT jannaschii."

RL Science 273:1058-1073(1996).

CC -1- SIMILARITY: BELONGS TO THE HELICASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC EMBL: U67484; AAB98279.1; -.

DR TIGR: M0294;

DR InterPro: IPR001410; DEAD.

DR InterPro: IPR001650; Helicase_C.

DR Pfam: PF00270; DEAD. 1.

DR Pfam: PF00271; helicase_C. 1.

DR SMART: SM00487; DEXDC. 1.

DR SMART: SM00490; HELIC_C. 1.

KW Hypothetical protein; Hydrolase; Helicase; ATP-binding;

KW Complete proteome.

FT NE BIND 56

FT SITE 181 184

FT SEQUENCE 841 AA; 97267 MW; 4C28A8A9459505 CRC64;

SQ SEQUENCE 841 AA; 97267 MW; 4C28A8A9459505 CRC64;

QY 360 EKLKGEIT 367

IIIIIIII

DB 323 EKLKGEIT 330

RESULT 4

MGR4_HUMAN STANDARD; PRT; 912 AA.

Query Match 1.8%; Score 8; DB 1; Length 841;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 360 EKLKGEIT 367

IIIIIIII

DB 323 EKLKGEIT 330

RESULT 4

MGR4_HUMAN STANDARD; PRT; 912 AA.

Query Match 1.8%; Score 8; DB 1; Length 841;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 360 EKLKGEIT 367

IIIIIIII

DB 323 EKLKGEIT 330

RESULT 4

MGR4_HUMAN STANDARD; PRT; 912 AA.

Query Match 1.8%; Score 8; DB 1; Length 841;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 360 EKLKGEIT 367

IIIIIIII

DB 323 EKLKGEIT 330

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 11:59:35 ; Search time 14 Seconds
(without alignments)

1315.392 Million cell updates/sec

Title: US-09-881-752A-212

Perfect score: 444
Sequence: 1 MKHFSVKRLGLSLVLT.....VRVANEYFKDTOSTVFLKP 444

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	2.0	636	RPB1_METJA	O60181 methanococ
2	8	1.8	160	RUVQ_CAMJE	Q9P1U8 campylobact
3	8	1.8	841	HELD_METJA	Q57742 methanococ
4	8	1.8	912	MGR4_HUMAN	Q14833 homo sapien
5	7	1.6	141	2SS8_HELAN	P23110 heilanthus
6	7	1.6	146	HBH2_XENTR	P08423 xenopus tro
7	7	1.6	146	HBH2_XENTR	P11342 xenopus tro
8	7	1.6	146	HBH2_XENTR	P11342 xenopus tro
9	7	1.6	172	PHCB_STNP7	P06539 synchococ
10	7	1.6	194	FM98_BCOLI	P21413 escherichia
11	7	1.6	203	Y461_PYRHO	O58214 pyrococcus
12	7	1.6	212	YTFB_BCOLI	P39310 escherichia
13	7	1.6	220	NHAB_PSECL	P27763 pseudomonas
14	7	1.6	223	Y770_METJA	O58180 methanococ
15	7	1.6	226	Y770_METJA	O58180 methanococ
16	7	1.6	233	LYSE_CORGL	P94633 corynebacte
17	7	1.6	245	UNG_MYCGE	P77343 mycoplasma
18	7	1.6	284	YKGD_BCOLI	P77373 escherichia
19	7	1.6	310	ISPH_CHLPN	O926P2 chlamydia p
20	7	1.6	334	SRK3_SPOLA	P42689 sponsigilla
21	7	1.6	346	BCH1_CHLVI	O50312 chlorobium
22	7	1.6	353	SOHB_HAEIN	P45015 haemophilus
23	7	1.6	357	MURG_CLOPE	O8X1Q1 clostridium
24	7	1.6	360	MRAY_RHILC	O98K80 rhizobium
25	7	1.6	370	RFL_MYCCA	P71496 mycoplasma
26	7	1.6	375	Y147_MYCGE	P47393 mycoplasma
27	7	1.6	376	ACTC_HALRO	P53461 halocynthia
28	7	1.6	377	ACTE_STRPU	P18499 strongyloce
29	7	1.6	377	Y147_MYCPN	P75585 mycoplasma
30	7	1.6	381	CDRA_SCHPO	O09184 schizosacch
31	7	1.6	386	METB_BCOLI	P00335 escherichia
32	7	1.6	396	YD61_MYCTU	Q10131 mycobacteri
33	7	1.6	415	CLPX_TREPA	O28303 archaeoglob

ALIGNMENTS

RESULT 1
ID RPB1_METJA STANDARD: PRT: 636 AA.
AC O60181:
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE DNA-directed RNA polymerase subunit B (EC 2.7.7.6).
GN RPB1 OR MJ1041.
OS Methanococcus jannaschii.
OC Archaea: Euryarchaeota: Methanococci: Methanococcales;
OC Methanocaldococcaceae: Methanocaldococcus.
OX NCBI_Taxid=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999: PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Nguyen D.,
RA Scott J.L., Geoghegan N.S.M., Weldman J.F., Fuhrmann J.L., Nuyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
*Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.*;
Science 273:1058-1073(1996).
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
SUBSTRATES. THE B ('B'+B' AND BETA) SUBUNITS HAVE BEEN IMPLICATED
IN DNA PROMOTER RECOGNITION AND ALSO IN NUCLEOTIDE BINDING.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
{RNA}(n).
CC -!- COFACTOR: ZINC.
CC -!- SIMILARITY: THE COMBINED B'+B' SUBUNITS ARE EQUIVALENT TO THE B
SUBUNITS OF THE EUKARYOTIC POLYMERASES I AND II AND TO EUKARYOTIC
BETA SUBUNIT.
CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC -----
CC EMBL: 067546; AAB99044.1; -
CC TIGR: MJ1041; -
CC InterPro: IPR001572: RNA_pol_B.
DR Pfam: PF00562: RNA_pol_B: 1
DR PROSITE: PS01166: RNA_POL_BETA: 1
KW Transferase; Transcription; DNA-directed RNA polymerase; Zinc;
KW Complete proteome; zinc-finger.
KW ZN_FING 551 572 POTENTIAL.

Search completed: November 18, 2002, 12:08:23
Job time : 13 secs

OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/050,516
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/828,855
FILING DATE: 31-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Becker, Cheryl L.
REGISTRATION NUMBER: 35,441
REFERENCE/DOCKET NUMBER: 6065.US.P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847/935-1729
TELEFAX: 847/938-2623
TELEX:
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. US20010010904A1e
US-09-050-516-44

Query Match 1.4%; Score 6; DB 10; Length 31;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 155 VAEER 160
Db 25 VAEER 30

RESULT 15
US-10-057-558-12
Sequence 12, Application US/10057558
Patent No. US20020164585A1
GENERAL INFORMATION:
APPLICANT: Chapman, Sean
APPLICANT: Dawson, William O.
APPLICANT: Donson, Jonathan
APPLICANT: Kumagai, Monto H.
APPLICANT: Lewandowski, Dennis J.
APPLICANT: Lindo, John A.
APPLICANT: Pogue, Gregory P.
APPLICANT: Shiyprasad, Shailaja
TITLE OF INVENTION: METHOD FOR ENHANCING RNA OR PROTEIN
TITLE OF INVENTION: PRODUCTION USING NON-NATIVE 5' UNTRANSLATED SEQUENCES IN
FILE REFERENCE: 008010137US08
CURRENT APPLICATION NUMBER: US/10/057,558
CURRENT FILING DATE: 2002-01-25
PRIOR APPLICATION NUMBER: 09/359,299
PRIOR FILING DATE: 1999-07-21
NUMBER OF SEQ ID NOS: 49
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 12
LENGTH: 32
TYPE: PRT
ORGANISM: Tobacco mosaic virus
US-10-057-558-12

Query Match 1.4%; Score 6; DB 9; Length 32;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 LIGLSS 14
Db 21 LIGLSS 26

GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aemica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/532,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 36290
LENGTH: 26
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL008632.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.4
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.3
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.4
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.7
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6.4
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.6
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4
OTHER INFORMATION: EST_HUMAN HT: BF304683.1, EVALUATE 4.00e-03
US-09-864-761-36290

Query Match 1.4%: Score 6; DB 10; Length 26;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
141 LINKED 146
|||||

Db 5 LINKED 10
RESULT 13
US-10-057-558-6
Sequence 6, Application US/10057558
Patent No. US20020164585A1
GENERAL INFORMATION:
APPLICANT: Chapman, Sean
APPLICANT: Dawson, William O.
APPLICANT: Donson, Jonathan
APPLICANT: Kumagai, Monto H.
APPLICANT: Lewandowski, Dennis J.
APPLICANT: Lindo, John A.
APPLICANT: Pogue, Gregory P.
APPLICANT: Shivprasad, Shailaja
TITLE OF INVENTION: METHOD FOR ENHANCING RNA OR PROTEIN
TITLE OF INVENTION: PRODUCTION USING NON-NATIVE 5' UNTRANSLATED SEQUENCES IN
FILE REFERENCE: 008010137US08
CURRENT APPLICATION NUMBER: US/10/057,558
PRIOR FILING DATE: 2002-01-25
PRIOR APPLICATION NUMBER: 09/359,299
PRIOR FILING DATE: 1999-07-21
NUMBER OF SEQ ID NOS: 49
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 6
LENGTH: 31
TYPE: PRT
ORGANISM: Tobacco mosaic virus
US-10-057-558-6

Query Match 1.4%: Score 6; DB 9; Length 31;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 LIGLSS 14
|||||
Db 19 LIGLSS 24

RESULT 14
US-09-050-516-44
Sequence 44, Application US/09050516
Patent No. US20010010904A1
GENERAL INFORMATION:
APPLICANT: BILLING-MEDEL, PATRICIA
APPLICANT: COHEN, MAURICE
APPLICANT: COLPITTS, TRACEY L.
APPLICANT: FRIEDMAN, PAULA N.
APPLICANT: GORDON, JULIAN
APPLICANT: GRANADOS, EDWARD N.
APPLICANT: HAYDEN, MARK
APPLICANT: HODGES, STEVEN C.
APPLICANT: KLAAS, MICHAEL R.
APPLICANT: KRATOCHVIL, JON D.
APPLICANT: ROBERTS-RAPP, LISA
APPLICANT: RUSSELL, JOHN C.
APPLICANT: STROUPE, STEPHEN D.
TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
TITLE OF INVENTION: FOR DETECTING DISEASES OF THE GASTROINTESTINAL
TITLE OF INVENTION: TRACT
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESS: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible

; TITLE OF INVENTION: No. US20020142428A1el kinases and uses thereof
 ; FILE REFERENCE: 35800/234862
 ; CURRENT APPLICATION NUMBER: US/09/862,027
 ; CURRENT FILING DATE: 2001-05-21
 ; PRIOR APPLICATION NUMBER: US 09/345,473
 ; PRIOR FILING DATE: 1999-06-30
 ; NUMBER OF SEQ ID NOS: 82
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 33
 ; LENGTH: 553
 ; TYPE: PRF
 ; ORGANISM: Arabidopsis thaliana
 ; US-09-862-027-33

Query Match 1.6%; Score 7; DB 10; Length 553;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 401 LVQNDIQ 407
 |||||
 Db 522 LVQNDIQ 528

RESULT 9
 ; US-09-815-242-10109
 ; Sequence 10109, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Karl L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 10109
 ; LENGTH: 572
 ; TYPE: PRF
 ; ORGANISM: Escherichia coli
 ; US-09-815-242-10109

Query Match 1.6%; Score 7; DB 10; Length 572;
 Best Local Similarity 100.0%; Pred. No. 74;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 PEROVA 156
 |||||
 Db 423 PEROVA 429

RESULT 10

; US-09-877-804-16
 ; Sequence 16, Application US/09877804
 ; Patent No. US20020061557A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Nikolic, Karoly
 ; APPLICANT: McFarland, Keith C.
 ; APPLICANT: Segalo, Deborah L.
 ; APPLICANT: Seaburg, Peter H.
 ; TITLE OF INVENTION: Glycoprotein Hormone Receptor Molecules
 ; FILE REFERENCE: P0576P1C2
 ; CURRENT APPLICATION NUMBER: US/09/877,804
 ; CURRENT FILING DATE: 2001-09-04
 ; PRIOR APPLICATION NUMBER: US 08/207,814
 ; PRIOR FILING DATE: 1994-03-07
 ; PRIOR APPLICATION NUMBER: US 07/781,153
 ; PRIOR FILING DATE: 1991-10-31
 ; PRIOR APPLICATION NUMBER: US 07/347,683
 ; PRIOR FILING DATE: 1989-05-05
 ; NUMBER OF SEQ ID NOS: 22
 ; SEQ ID NO 16
 ; LENGTH: 10
 ; TYPE: PRF
 ; ORGANISM: Artificial sequence
 ; FEATURE:
 ; OTHER INFORMATION: fragment
 ; US-09-877-804-16

Query Match 1.4%; Score 6; DB 10; Length 10;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 ESVTLK 39
 |||||
 Db 4 ESVTLK 9

RESULT 11
 ; US-09-852-555-3
 ; Sequence 3, Application US/09852555
 ; Patent No. US20020034751A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Yuen, Kwok-Yung
 ; APPLICANT: Cao, Liang
 ; TITLE OF INVENTION: Aspergillus Fumigatus Antigenic Protein 1
 ; FILE REFERENCE: 609920-600015
 ; CURRENT APPLICATION NUMBER: US/09/852,555
 ; CURRENT FILING DATE: 2001-05-10
 ; PRIOR APPLICATION NUMBER: US 60/203,322
 ; PRIOR FILING DATE: 2000-05-10
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 3
 ; LENGTH: 17
 ; TYPE: PRF
 ; ORGANISM: Aspergillus fumigatus
 ; FEATURE:
 ; NAME/KEY: SIGNAL
 ; LOCATION: (1)..(17)
 ; US-09-852-555-3

Query Match 1.4%; Score 6; DB 10; Length 17;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 LVTVLG 21
 |||||
 Db 6 LVTVLG 11

RESULT 12
 ; US-09-864-761-36290
 ; Sequence 36290, Application US/09864761
 ; Patent No. US20020048763A1

```
; APPLICANT: zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11084
; LENGTH: 415
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-815-242-11084
```

```
Query Match      1.6%; Score 7; DB 10; Length 415;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 357 ALLEKIK 363
DB 258 ALLEKIK 264
```

```
RESULT 6
US-09-815-242-13781
; Sequence 13781, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
```

```
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13781
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-09-815-242-13781
```

```
Query Match      1.6%; Score 7; DB 10; Length 433;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 438 TTVFLKP 444
DB 131 TTVFLKP 137
```

```
RESULT 7
US-09-815-242-12079
; Sequence 12079, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12079
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-12079
```

```
Query Match      1.6%; Score 7; DB 10; Length 474;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 219 VLVVGDV 225
DB 13 VLVVGDV 19
```

```
RESULT 8
US-09-862-027-33
; Sequence 33, Application US/09862027
; Patent No. US20020142428A1
; GENERAL INFORMATION:
; APPLICANT: Hodge, Martin R.
```

QY 181 VYHPYHMTPIGEMDIONMTLKDIKKFSHYQPKMAIVLVGVDSQKVFELSKKHES 240
| | | | |
Db 181 VYHPYHMTPIGEMDIONMTLKDIKKFSHYQPKMAIVLVGVDSQKVFELSKKHES 240
QY 241 LKNIDKAIIPPYKPKKODGARTAVVHKDGVHLEWALGKYPAFKHKDQVALDALSR 300
| | | | |
Db 241 LKNIDKAIIPPYKPKKODGARTAVVHKDGVHLEWALGKYPAFKHKDQVALDALSR 300
QY 301 LGEKSSWLOSELVDKRLKLSQAFSHNMQLODESVFLFIAGNPNVKAALOKETVALLE 360
| | | | |
Db 301 LGEKSSWLOSELVDKRLKLSQAFSHNMQLODESVFLFIAGNPNVKAALOKETVALLE 360
QY 361 KLRKGEITQALDKLTKINQKADFTISNLESSSDVAGLFADYLVQNDIGLTDYQROFLDLK 420
| | | | |
Db 361 KLRKGEITQALDKLTKINQKADFTISNLESSSDVAGLFADYLVQNDIGLTDYQROFLDLK 420
QY 421 VSDLYRVANEFKDTOSTTYVFLKP 444
| | | | |
Db 421 VSDLYRVANEFKDTOSTTYVFLKP 444

RESULT 2

US-09-817-464-2
; Sequence 2, Application US/09817464
; Patent No. US20020127638A1
; GENERAL INFORMATION:
; APPLICANT: Flör, Peter J.
; APPLICANT: Kuhn, Rainer
; APPLICANT: Lindaur, Kristen
; APPLICANT: Puttner, Irene
; APPLICANT: Knopfel, Thomas
; TITLE OF INVENTION: Human Metabotropic Glutamate Receptor Subtypes (HMR4,
; FILE REFERENCE: 4-19679/4/PCT
; CURRENT APPLICATION NUMBER: US/09/817,464
; EARLIER FILING DATE: 2001-03-26
; EARLIER APPLICATION NUMBER: US/08/617,785
; EARLIER FILING DATE: 1996-03-19
; EARLIER APPLICATION NUMBER: EPO 9416553.7
; EARLIER FILING DATE: 1994-08-19
; EARLIER APPLICATION NUMBER: EPO 93810663.0
; EARLIER FILING DATE: 1993-09-20
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 912
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-817-464-2

Query Match 1.8%; Score 8; DB 10; Length 912;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 92 KAGEFDKI 99
| | | | |
Db 261 KAGEFDKI 268

RESULT 3

US-09-796-100-5
; Sequence 5, Application US/09796100
; Patent No. US20020076784A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 40322, A No. US20020076784A1 Human Dynamn
; FILE REFERENCE: 35800/209283
; CURRENT APPLICATION NUMBER: US/09/796,100
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/185,503
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5
; LENGTH: 298
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Pfam consensus sequence
US-09-796-100-5

Query Match 1.6%; Score 7; DB 10; Length 298;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 348 AEALQKE 354
| | | | |
Db 34 AEALQKE 40

RESULT 4

US-09-815-242-10427
; Sequence 10427, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10427
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10427

Query Match 1.6%; Score 7; DB 10; Length 386;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 438 TTVFLKP 444
| | | | |
Db 84 TTVFLKP 90

RESULT 5

US-09-815-242-11084
; Sequence 11084, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 12:04:30 ; Search time 12 Seconds

(without alignments)
557.253 Million cell updates/sec

Title: US-09-881-752a-212

Perfect score: 444
Sequence: 1 MKHFSVKRLGLISVLLVTL.....VRVANEYFKDTOSTYFLKP 444

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 97044 seqs, 15060890 residues

Word size : 0

Total number of hits satisfying chosen parameters: 97044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications_AA.*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	444	100.0	444	10	US-09-881-752a-212
2	8	1.8	912	10	US-09-817-464-2
3	7	1.6	298	10	US-09-796-100-5
4	7	1.6	386	10	US-09-815-242-10427
5	7	1.6	415	10	US-09-815-242-11084
6	7	1.6	433	10	US-09-815-242-13781
7	7	1.6	474	10	US-09-815-242-12079
8	7	1.6	553	10	US-09-862-027-33
9	7	1.6	572	10	US-09-815-242-10109
10	6	1.4	10	10	US-09-877-804-16
11	6	1.4	17	10	US-09-852-555-3
12	6	1.4	26	10	US-09-864-761-36290
13	6	1.4	31	10	US-10-057-558-6
14	6	1.4	31	10	US-09-050-516-44
15	6	1.4	32	9	US-10-057-558-12
16	6	1.4	33	9	US-10-057-558-4
17	6	1.4	33	10	US-09-864-761-44547
18	6	1.4	34	10	US-09-864-761-47905
19	6	1.4	40	9	US-10-057-558-10

20	6	1.4	42	9	US-09-984-245-289	Sequence 289, App
21	6	1.4	43	10	US-09-864-761-33835	Sequence 33835, A
22	6	1.4	48	10	US-09-864-761-44620	Sequence 44620, A
23	6	1.4	56	10	US-09-864-761-44297	Sequence 44297, A
24	6	1.4	56	10	US-09-945-173-9	Sequence 9, Appli
25	6	1.4	60	9	US-10-057-558-8	Sequence 8, Appli
26	6	1.4	62	10	US-09-864-761-34486	Sequence 34486, A
27	6	1.4	64	10	US-09-864-761-48364	Sequence 48364, A
28	6	1.4	68	10	US-09-764-877-1394	Sequence 1394, App
29	6	1.4	73	10	US-09-216-393-41	Sequence 41, Appl
30	6	1.4	73	10	US-09-864-761-48385	Sequence 48385, A
31	6	1.4	81	10	US-09-923-301-1421	Sequence 1421, Ap
32	6	1.4	85	10	US-09-815-242-10319	Sequence 10319, A
33	6	1.4	85	10	US-09-815-242-11140	Sequence 11140, A
34	6	1.4	85	10	US-09-815-242-13771	Sequence 13771, A
35	6	1.4	87	10	US-09-815-242-11763	Sequence 11763, A
36	6	1.4	90	10	US-09-815-242-5313	Sequence 5313, Ap
37	6	1.4	90	10	US-09-815-242-12601	Sequence 12601, A
38	6	1.4	90	10	US-09-815-242-12756	Sequence 12756, A
39	6	1.4	90	10	US-09-815-242-13073	Sequence 13073, A
40	6	1.4	91	10	US-09-815-242-5022	Sequence 5022, Ap
41	6	1.4	91	10	US-09-815-242-10688	Sequence 10688, A
42	6	1.4	91	10	US-09-815-242-13319	Sequence 13319, A
43	6	1.4	94	10	US-09-864-761-48596	Sequence 48596, A
44	6	1.4	97	10	US-09-864-761-40638	Sequence 40638, A
45	6	1.4	97	10	US-09-867-550-356	Sequence 356, App

ALIGNMENTS

RESULT 1
US-09-881-752a-212
; Sequence 212, Application US/09881752a
; Patent No. US20020115078a1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020115078a1 Helicobacter Polypeptides in t
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 212
; LENGTH: 444
; TYPE: PRT
; ORGANISM: Helicobacter pylori

Query Match 100.0%; Score 444; DB 10; Length 444;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 444; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKHFSVKRLGLISVLLVTLGASMAHOSYLPKHESVTLKNGLOVSVPLENKTGVEVDV 60
DB 1 MKHFSVKRLGLISVLLVTLGASMAHOSYLPKHESVTLKNGLOVSVPLENKTGVEVDV 60
QY 61 LKVGSRNEMTGKSGIAHMLEHLNFKSTNKLKAGEPDKYKRGVSNASTSPDITRYFT 120
DB 61 LKVGSRNEMTGKSGIAHMLEHLNFKSTNKLKAGEPDKYKRGVSNASTSPDITRYFT 120
QY 121 KTSQANDLSLELPFAETMGSLNKEDEFLPEROVAAEBRRRWRDNPGLMTRFRFNAY 180
DB 121 KTSQANDLSLELPFAETMGSLNKEDEFLPEROVAAEBRRRWRDNPGLMTRFRFNAY 180